

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/148189>

Please be advised that this information was generated on 2017-12-05 and may be subject to change.

2739

17

Aspecten van begrijpend lezen in het vierde leerjaar van het gewoon lager onderwijs



C.A.J. Aarnoutse

ASPECTEN VAN BEGRIJPEND LEZEN IN HET VIERDE LEERJAAR VAN HET GEWOON LAGER ONDERWIJS

Voor Thea, Rob en Bea

Voorwoord	XII
I. <u>Inleiding</u>	1
1.1. Voorafgaand onderzoek	1
1.2. Aanleiding en doel van het onderzoek	4
1.3. Overzicht van het rapport	10
II. <u>Begrijpend lezen</u>	12
2.1. Inleiding	12
2.2. Twee benaderingen van begrijpend lezen	15
2.2.1. Het model van Kintsch en Van Dijk	17
2.2.2. De schematheorie van Rumelhart e.a.	24
2.3. Inferenties en inferentieprocessen	32
2.4. Selectie en specificatie van de leestaken	40
2.4.1. Selectie van de taken	40
2.4.2. Specificatie van de taken	42
- Het afleiden van de hoofdgedachte	43
- Het afleiden van de bedoeling	45
- Het afleiden van logisch geldige informatie (redeneren)	47
III. <u>Aspecten van begrijpend leesonderwijs</u>	51
3.1. Programma's voor begrijpend lezen en hun effect	51
3.1.1. Programma's voor begrijpend lezen in de Verenigde Staten	52
3.1.2. Programma's voor begrijpend lezen in Nederland	62
3.1.3. Het effect van strategieën en experimentele programma's	67
3.1.4. Opzet van de experimentele programma's	73

3.2. Vaardigheden op het gebied van begrijpend lezen	76
3.2.1. Vaardigheden van begrijpend lezen	77
3.2.2. Onderzoek naar vaardigheden van begrijpend lezen	81
3.3. Methoden en instrumenten voor het meten van begrijpend lezen	90
3.3.1. Methoden voor het meten van begrijpend lezen	91
3.3.2. Instrumenten voor het meten van begrijpend lezen	96
- Taakgerichte tests	100
- Cloze tests	106
 IV. <u>Opzet en methode van het onderzoek</u>	115
4.1. Inleiding	115
4.2. Vraagstellingen	116
4.3. Opzet van het onderzoek	117
4.4. Ontwikkeling van de tests	123
4.4.1. De Hoofdgedachtetest	123
- Taak en kenmerken van de teksten	123
- De ontwikkeling van de test	125
4.4.2. De Bedoelingtest	127
- Taak en kenmerken van de teksten	127
- De ontwikkeling van de test	129
4.4.3. De Redeneertest	132
- Taak en kenmerken van de teksten	132
- De ontwikkeling van de test	136
4.4.4. De Free Response Cloze Test	139
- Achtergrond	139
- De ontwikkeling van de test	140
4.4.5. De Multiple Choice Cloze Test	142
- Achtergrond	142
- De ontwikkeling van de test	143

4.5. Ontwikkeling van de programma's	147
4.5.1. Ontwikkeling van het programma Hoofdgedachte	147
- Doel en opzet van het programma	147
- Ontwikkeling van versie 1 en 2	149
- Definitieve versie van het programma	153
4.5.2. Ontwikkeling van het programma Bedoeling	153
- Doel en opzet van het programma	154
- Ontwikkeling van versie 1 en 2	155
- Definitieve versie van het programma	158
4.5.3. Ontwikkeling van het programma Redeneren	160
- Doel en opzet van het programma	160
- Ontwikkeling van versie 1 en 2	162
- Definitieve versie van het programma	165
4.6. Selectie van de scholen en vergelijking van de experimentele en controlescholen	167
4.6.1. Selectie van de scholen	167
4.6.2. Vergelijking van de experimentele en controle- scholen	170
4.7. Dataverzameling	173
4.7.1. Leerlinggegevens	174
4.7.2. Testafnames	174
- Tijdstippen van afname	174
- Testinstructeurs	177
4.7.3. Verloop van de testafnames	178
- Omstandigheden waaronder de tests zijn afgenomen	179
- Op- en aanmerkingen over de tests en het testmateriaal	180
- Afwezigheid van leerlingen	180
- Maximale testtijd	181
4.8. Voorbereiding en uitvoering van het experiment	183
4.8.1. Selectie van het controleprogramma	183
4.8.2. Selectie van de instructeurs	184
4.8.3. Verdeling van de instructeurs over de scholen en de programma's	185

4.8.4.	Vorbereiding van de instructeurs	186
4.8.5.	Vorbereiding van de leerkrachten	187
4.8.6.	Verdeling van de leerlingen over de programma's	188
4.8.7.	Vorbereiding en uitvoering van de lessen	191
4.8.8.	Evaluatie van de lessen en programma's	192
V.	<u>Analyse van de gegevens en resultaten</u>	195
5.1.	Het descriptieve gedeelte van het onderzoek	195
5.1.1.	Vraagstellingen en verwachtingen	195
5.1.2.	De prestaties van de leerlingen op de ontwikkelde tests bij de voormeting	197
	- Frequentieverdelingen	197
	- Verschillen in prestaties tussen sociale milieus	206
	- Verschillen in prestaties tussen scholen	209
5.1.3.	De betrouwbaarheid van de ontwikkelde tests	212
	- Opzet	212
	- Resultaten	216
	- Scoringswijzen van de FRCT en beoordelaar-betrouwbaarheid	223
5.1.4.	De validiteit van de ontwikkelde tests	225
	- Inleiding	225
	- Homogeniteit van de tests	226
	- Samenhang tussen de vijf tests	230
	- Samenhang met technisch lezen	236
	- Samenhang met begrijpend lezen	238
	- Samenhang met woordenschat	241
	- Samenhang met non-verbale intelligentie	243
	- Samenhang met het praktijkoordeel	246
	- Variantiecomponenten binnen de vijf tests	252
5.1.5.	De verandering in de prestaties op de ontwikkelde tests	263
	- Inleiding	263
	- Verandering in de groep als geheel	264
	- Individuele verschillen in verandering	268

- Factoren die samenhangen met individuele verschillen in verandering	270
5.1.6. Samenvatting	273
5.2. Het experimentele gedeelte van het onderzoek	279
5.2.1. Hypothesen	279
5.2.2. Data	281
5.2.3. Analyse van de gegevens binnen de experimentele scholen	285
- Analyseprocedure	285
- Toetsing van de hypothesen: nameting	287
- Toetsing van de hypothesen: retentiemeting	291
5.2.4. Analyse van de gegevens van de experimentele en controlescholen samen	295
- Analyseprocedure	295
- Toetsing van de hypothesen: nameting en retentiemeting	296
5.2.5. Samenvatting	301
VI. <u>Slotbeschouwingen</u>	305
6.1. Resultaten	305
6.2. Beperkingen van het onderzoek	308
6.3. Verder onderzoek	313
 Summary	 316
 Literatuur	 322
 Bijlagen	 337
 Curriculum Vitae	 373

VOORWOORD

Het onderzoek waarvan in dit proefschrift verslag wordt gedaan, werd uitgevoerd op de vakgroep Interdisciplinaire Studierichting Onderwijskunde van de Katholieke Universiteit te Nijmegen in de periode van juli 1977 tot maart 1980. De Stichting voor Onderzoek van het Onderwijs heeft door het verlenen van subsidie (S.V.O.-project 0298) de realisering van dit onderzoek mogelijk gemaakt.

Graag betuig ik mijn dank aan allen die aan het onderzoek hebben meegewerkt. Mijn collega's drs. A. Buitenhuis, drs. W. Koopman, mevr. drs. M. Poos-Akerboom en drs. G. Seegers dank ik voor hun inzet en deskundigheid waarmee zij samen met mij aan het onderzoek of een deel ervan hebben gewerkt. Een speciaal woord van dank komt toe aan drs. J. van Leeuwe en drs. M. Voeten die tijdens het onderzoek en bij de vormgeving van het manuscript waardevolle adviezen gaven inzake statistische en methodologische kwesties. Prof. dr. A.J.A. Felling dank ik voor zijn opmerkingen bij de eindversie van het manuscript.

Veel dank ben ik ook verschuldigd aan de hoofden, leerkrachten én leerlingen van de scholen die belangeloos hun medewerking verleenden aan het onderzoek, aan de inspecteurs van de inspectie Nijmegen, Wijchen en Elst, aan de test- en programma-instructeurs, aan de assistenten en leerkrachten die geholpen hebben bij de ontwikkeling van de programma's, aan de samensteller van de vragenlijst en aan de assistent dataverwerking. Tenslotte dank ik mevr. A. van Loveren voor de goede zorg en de grote nauwgezetheid waarmee zij het manuscript heeft getypt.

Nijmegen, maart 1982

Cor Aarnoutse

Het onderzoek waarvan hier verslag wordt gedaan, past binnen een bepaald kader en is een vervolg op eerder verricht onderzoek. Dat is de reden dat in par. 1.1. aandacht wordt besteed aan het onderzoeksplan van de themagroep Lees-Taalonderwijs en aan twee eerder uitgevoerde onderzoeken. In par. 1.2. volgt een beschrijving van de aanleiding en het doel van het onderhavige onderzoek. Par. 1.3. geeft een overzicht van het verslag.

1.1. Voorafgaand onderzoek

Sedert een aantal jaren functioneert binnen de vakgroep Interdisciplinaire Studierichting Onderwijskunde te Nijmegen de themagroep Lees-Taalonderwijs. Deze groep heeft in 1972 een "Plan-in-Hoofddijnen" opgesteld met betrekking tot onderzoek op het gebied van lees-taalonderwijs. Dit plan geeft in grote lijnen aan welk onderzoek in de loop der jaren op het betreffende gebied wordt uitgevoerd. In dit plan wordt het uiteindelijke doel van het gehele onderzoek als volgt omschreven:

- a. het formuleren van doelstellingen voor het leesonderwijs in de lagere school;
- b. het ontwikkelen van didactische en diagnostische werkwijzen gericht op het bereiken van deze doelstellingen;
- c. het ontwerpen in hoofddijnen van een curriculum voor het leesonderwijs in de lagere school.

Op basis van dit plan werd van september 1974 tot juli 1976 een onderzoek uitgevoerd naar de doelstellingen van een aantal systematisch ontwikkelde leesprogramma's voor de lagere school. Dit onderzoek was op twee doelen gericht, nl. (a) nagaan - voor zover mogelijk - op welke wijze en op grond van welke theorieën, concepten en wetenschappelijke gegevens de doelstellingen van een aantal systematisch ontwikkelde leesprogramma's zijn vastgesteld, geformuleerd en geordend, (b) ontwikkelen van een voorlopige lijst van doelstellingen van het leesonderwijs voor de lagere school.

In de eerste fase van het onderzoek werd een literatuurstudie verricht naar die aspecten van lezen en leesonderwijs, die direct of indirect betrekking hebben op de problematiek van de doelstellingen van het leesonderwijs. Uit deze studie bleek dat er nog geen coherente leestheorie is ontwik-

keld en dat nog veel onderzoek op het gebied van de doelstellingen van het leesonderwijs moet worden verricht. Voor meer informatie verwijzen we naar Deelrapport 1 "Enkele aspecten van lezen en leesonderwijs" (Aarnoutse et al., 1975).

In de tweede fase van het onderzoek werden zes Amerikaanse en twee Nederlandse leesprogramma's bestudeerd en geanalyseerd met behulp van een daartoe ontwikkeld analyseschema. Voor een beschrijving van deze programma's en voor de analyse van de verschillende soorten van doelstellingen verwijzen we naar Deelrapport 2 "Doelstellingen van leesonderwijs" (Aarnoutse et al., 1976). De belangrijkste conclusie van deze analyse was dat de indeling, bepaling en sequentiëring van de doelen van de verschillende programma's nog voor een groot deel berust op voorwetenschappelijke kennis en ervaring.

Op basis van de literatuurstudie, de analyse van de doelen van de verschillende leesprogramma's en een aantal doelstellingenoverzichten werd vervolgens een voorlopige lijst van doelstellingen van het leesonderwijs voor de lagere school ontwikkeld. Deze lijst bevat vier categorieën van doelen, nl. decoderen, begrijpend lezen, verwerven van informatie door lezen en houding ten aanzien van lezen en leesmateriaal. Elke categorie is onderverdeeld in drie subcategorieën. In de lijst worden de categorieën zo goed mogelijk omschreven en verantwoord. Elke subcategorie bevat een aantal doelstellingen die tenminste in termen van gedrag zijn geformuleerd. In Deelrapport 2 wordt de voorlopige lijst van doelstellingen weergegeven. Deze lijst vormde een basis voor empirisch onderzoek.

Van juli 1976 tot juni 1977 werd een empirisch onderzoek uitgevoerd dat onder andere tot doel had (a) de prestaties van de leerlingen van de hoogste leerjaren van de lagere school vast te stellen met betrekking tot een aantal doelen op het gebied van begrijpend lezen, (b) de realiteitswaarde van het onderscheid tussen het begrijpen van de letterlijke betekenis en impliciete betekenis na te gaan.

Wat betreft de keuze van de doelstellingen werd uitgegaan van de voorlopige lijst van doelstellingen en van het model van Levelt en Kempen (1976) waarin onder andere onderscheid wordt gemaakt tussen lexicale, syntactische en textuele vaardigheden. Voor elk vaardigheidstype werd één doelstelling geselecteerd (woord-, zins- en tekstniveau). Bovendien werden twee doelstellingen gekozen, die betrekking hadden op het onderscheid tussen het begrijpen van de letterlijke en impliciete betekenis. De volgende doelstellingen

werden gekozen:

- de leerling kan synoniemen identificeren (woordniveau);
- de leerling kan bij een actieve zin de bijbehorende passieve zin met dezelfde betekenis vinden en omgekeerd (zinsniveau);
- de leerling kan de hoofdgedachte van een korte tekst identificeren (tekstniveau, expliciet);
- de leerling kan de hoofdgedachte uit een korte tekst afleiden (tekstniveau, impliciet);
- de leerling kan de hoofdgedachte uit een tekst van twee alinea's afleiden (tekstniveau, impliciet).

Voor elk van de bovengenoemde doelstellingen werden tests ontwikkeld. Bovendien werden twee cloze^x tests geconstrueerd. Deze leestests bestaan uit teksten van ongeveer 350 woorden waarin elk zevende woord is weggelaten. Op de open plaatsen in de tekst dienen de leerlingen de weggelaten woorden in te vullen. Het aantal goed ingevulde woorden vormt een maat voor tekstbegrip.

De vraagstellingen van dit onderzoek hadden grotendeels betrekking op de kwaliteit (met name de homogeniteit) van de ontwikkelde tests, de prestaties van de leerlingen op deze tests en de relaties tussen de totaalscores op de tests. Twee onderzoeksvragen hadden betrekking op de indeling van de doelstellingen. Bij één onderzoeksvraag werd gevraagd naar het verschil in prestatie tussen een test die betrekking had op het identificeren van de hoofdgedachte en een bijna identieke test die betrekking had op het afleiden van de hoofdgedachte (het begrijpen van de letterlijke betekenis versus het begrijpen van de impliciete betekenis). Bij een andere onderzoeksvraag werd gevraagd in hoeverre de data aanwijzingen bevatten dat de indeling van de doelstellingen in vaardigheden op woord-, zins- en tekstniveau reëel is.

In totaal waren vier grote lagere scholen (met dubbelklassen) in Nijmegen bij het onderzoek betrokken. De vijf doelstellinggerichte tests werden bij ± 200 leerlingen van het 5e en 6e leerjaar afgenomen evenals de cloze tests, de Stilleestest van Bakker (1972) en de Woordenschattest van Stijnen (1975). Het merendeel van deze tests werd ook in het 3e en 4e leerjaar van deze scholen afgenomen.

Uit het onderzoek bleek dat de homogeniteit van met name de drie hoofdgedachtetests voldeed aan de eisen die algemeen gesteld worden. De tests op

^x De term "cloze" is een samentrekking van "closure", een principe uit de Gestaltpsychologie.

woord- en zinsniveau waren minder homogeen dan de hoofdgedachtetests. Verder bleek dat er een verschil bestond tussen het identificeren en afleiden van de hoofdgedachte van een tekst. Dit verschil was echter niet van dien aard dat hierdoor de indeling in het begrijpen van de letterlijke en impliciete betekenis in sterke mate werd ondersteund. Tenslotte bleek dat de indeling van de doelstellingen in vaardigheden op woord-, zins- en tekstniveau tot op zekere hoogte in de factoranalyse was terug te vinden.

Voor meer informatie omtrent het doel, de opzet en de resultaten van dit onderzoek verwijzen we naar Deelrapport 3 "Een onderzoek naar de mate waarin een aantal doelstellingen van begrijpend lezen wordt bereikt" (Aarnoutse et al., 1977). In dit rapport wordt ook verslag gedaan van een literatuurstudie die betrekking had op het proces van begrijpend lezen, vaardigheden op het gebied van begrijpend lezen, buitenlandse tests voor begrijpend lezen en psychometrische methoden en technieken. Bovendien worden in dit rapport de belangrijkste implicaties van de literatuurstudie en van het empirisch onderzoek vermeld. Een belangrijke implicatie was dat het theoretisch niet zinvol is onderzoek te doen naar de realiteitswaarde van het onderscheid tussen het begrijpen van letterlijke en impliciete informatie. Een andere belangrijke implicatie voor het onderhavige onderzoek was dat begrijpend lezen primair op tekstniveau onderzocht moet worden en dat hierbij het accent moet liggen op die leestaken, waarbij leerlingen informatie uit een tekst moeten afleiden (in het vervolg "inferentiële leestaken" genoemd).

1.2. Aanleiding en doel van het onderzoek

Hoewel het aanvankelijk in de bedoeling lag een soortgelijk descriptief onderzoek te doen als hierboven beschreven, werd later besloten een meer praktisch gericht onderzoek uit te voeren met als doel een bijdrage te leveren aan de curriculumontwikkeling op het gebied van begrijpend lezen. Dit ondanks het feit dat vanaf het begin duidelijk was dat een dergelijk onderzoek riskant is gezien de beperkte kennis van het proces van begrijpend lezen en van het begrijpend leesonderwijs.

De aanleiding om het onderhavige onderzoek uit te voeren, vormt vooral het Wisconsin Design for Reading Skill Development. Voor een goed begrip van het doel van het onderzoek volgt nu een korte beschrijving van dit leesprogramma. Voor uitvoerige informatie verwijzen we naar Deelrapport 2 (Aarnoutse et al., 1976).

Het Wisconsin Design is ontwikkeld door een projectgroep onder leiding van Otto (1977) aan het Wisconsin Research and Development Center for Cognitive Learning van de universiteit van Wisconsin. Het is een leesprogramma dat tot doel heeft leerkrachten van de kleuter- en lagere school (K-6) te helpen bij de invoering van een systeem van individueel geleid onderwijs (Individually Guided Education) ten behoeve van de ontwikkeling van essentiële leesvaardigheden bij kinderen van 6 tot 12 jaar. Toepassing van het systeem van individueel geleid onderwijs met behulp van het Wisconsin Design houdt in, dat het niveau van de leesvaardigheidsontwikkeling van elke leerling wordt vastgesteld, dat de leerlingen worden verdeeld in groepen die dezelfde vaardigheid of combinatie van vaardigheden trachten te verwerven, dat er mogelijkheden worden gecreëerd voor individuele hulp en voor zelfstandig werken of studie, dat het niveau van de leesvaardigheidsontwikkeling van elke leerling steeds opnieuw wordt vastgesteld en dat op basis hiervan telkens nieuwe groepen worden geformeerd.

Het Wisconsin Design bestaat uit een aantal componenten. De "Outline of Reading Skills" en de "Statement of Skills and Objectives" vormen de basis van het programma, omdat hierin de belangrijkste leesvaardigheden en doelstellingen worden vermeld die de leerlingen in de loop van het programma moeten verwerven respectievelijk bereiken. Een tweede component vormen de doelstellingerichte tests en de informele evaluatie-instrumenten voor de meting en beoordeling van de leesprestaties. Een derde component vormen de bronnenboeken voor de leerkracht (Teacher's Resource Files) waarin per leesvaardigheid en doelstelling wordt aangegeven welke materialen en oefeningen van bestaande leesprogramma's gebruikt kunnen worden. Bovendien bevatten de bronnenboeken ook veel zelfontwikkelde materialen, die per doelstelling aanwijzingen en suggesties bevatten voor mogelijke onderwijsactiviteiten en -procedures. De vierde component omvat een aantal boeken waarin uitvoerige richtlijnen worden gegeven voor de invoering en planning van het Design. In deze publicaties wordt bovendien een beschrijving gegeven van de uitgangspunten, doelstellingen en ontwikkeling van het Design.

De "Outline of Reading Skills" vormt het fundament van het gehele Design. In dit ontwerp worden zes gebieden van lezen onderscheiden nl. decoderen (word attack), begrijpend lezen, studievaardigheden, zelfstandig lezen, interpretatief lezen en creatief lezen. In de "Statement of Skills and Objectives" worden behalve de 198 vaardigheden ook de meer concrete doelstellingen vermeld.

Met uitzondering van decoderen worden alle gebieden in een aantal componenten (strands) onderscheiden. Elke component bevat een aantal vaardigheden, die in niveaus (A-G) zijn gerangschikt en die opklimmen in moeilijkheidsgraad. Deze niveaus komen over het algemeen overeen met de traditionele klassenniveaus (K-6). De vaardigheden en doelen die binnen het gebied "comprehension" worden onderscheiden, worden in de experimentele editie van 1974 in zes componenten verdeeld nl. hoofdgedachte, volgorde, redeneren, detail, contextuele aanwijzingen en affixen. Elke component bevat een aantal vaardigheden en doelen, die nauw met elkaar verwant zijn en die in een hiërarchische volgorde zijn geplaatst. Ter illustratie volgen hier de vaardigheden voor de component hoofdgedachte:

Level A: Identifies a topic: pictures.

Level B: Identifies a topic: paragraphs (mondeling aangeboden).

Level C: Identifies a topic: paragraphs (schriftelijk aangeboden).

Level D: Identifies a topic sentence.

Level E: Identifies a main idea: paragraph.

Level F: Identifies a main idea: two paragraphs (vgl. Otto, 1977, pag. 208-213).

Opgemerkt moet worden dat de bovenstaande vaardigheden verder worden geconcretiseerd in de vorm van doelen waarbij het gedragsaspect en gedeeltelijk ook het inhoudsaspect worden weergegeven.

Uit deze korte beschrijving kan worden afgeleid dat het Wisconsin Design een flexibel programma is, gebaseerd op een aantal belangrijke didactische principes. Het programma is op een systematische wijze ontwikkeld en in een groot aantal scholen beproefd. Het onderzoek heeft vooral betrekking op decoderen, begrijpend lezen en studievaardigheden en op het functioneren van het systeem. Uit publicaties van Van Blaricom en White (1975; 1976) en Otto (1977) blijkt dat bij het begin van het onderhavige onderzoek het onderdeel voor begrijpend lezen nog volop in ontwikkeling was en er nog geen duidelijkheid bestond omtrent het effect hiervan op leesprestaties.

Het Wisconsin Design is gebaseerd op de "skill-approach" of de "subskill-approach". Het meest karakteristieke van deze benadering is dat - in tegenstelling tot de "holistic approach" waarvan Goodman (1968) en Smith (1978) voorstanders zijn - wordt uitgegaan van een reeks van vaardigheden op het gebied van lezen en dat in het leesonderwijs systematisch en vaak afzonderlijk aandacht wordt besteed aan de ontwikkeling van de verschillende vaardigheden.

De onderliggende assumptie is dat bij het leren van een complex proces zoals lezen het leerproces wordt vergemakkelijkt als het geheel wordt verdeeld in componenten, die ieder afzonderlijk worden geleerd waarna deze worden opgenomen in een integraal geheel (Samuels, 1976).

Een voordeel van deze benadering is dat het (begrijpend) leesonderwijs gericht en systematisch kan verlopen. De leerkracht kan bij deze vorm van onderwijs zijn aandacht richten op één bepaalde vaardigheid en de leerlingen leren hoe een dergelijke vaardigheid kan worden verworven. "Teachers cannot teach comprehension in the abstract any more than they can teach students intelligence or happiness..." (Athey, 1977, pag. 240). Analyse van een aantal Nederlandse leesprogramma's wijst er op dat het begrijpend leesonderwijs voor een deel bestaat uit de toepassing van een aantal leesvormen zoals klas-sikaal lezen, groepslezen, kringlezen en dergelijke zonder dat voldoende duidelijk is welke doelen of vaardigheden precies met deze leesvormen worden nagestreefd of ontwikkeld. Het lijkt erop alsof de didactische werkvormen eerder doel dan middel zijn en weinig gericht op het leren beheersen van een aantal belangrijke vaardigheden op het gebied van het verwerken van schriftelijke informatie. Een ander opvallend verschijnsel is de geringe aandacht voor het leren van strategieën die nodig zijn bij het verwerven van vooral complexe leesvaardigheden. Het begrijpend leesonderwijs lijkt voor een deel te bestaan uit het lezen van teksten en het beantwoorden van (de vele) vragen zonder dat er voldoende aandacht wordt besteed aan de vraag welke strategie of procedure gehanteerd kan worden bij de beantwoording van deze vragen. Een tweede voordeel van deze benadering is dat - zoals in het Wisconsin Design gebeurt - het begrijpend leesonderwijs kan worden gedifferentieerd. Een derde voordeel van de "skill-approach" bestaat hierin dat met behulp van specifieke tests kan worden nagegaan of en in hoeverre leerlingen een bepaalde vaardigheid beheersen. Bovendien kunnen met deze tests tekorten in de verschillende programma-onderdelen of in de onderwijs- en leersituaties worden opgespoord.

Naast een aantal voordelen kent deze benadering ook een aantal niet te onderschatten problemen. Zo maken de voorstanders van de "skill-approach" niet voldoende duidelijk wat bedoeld wordt met de term "skills" en baseren ze de keuze van de vaardigheden niet of in onvoldoende mate op een theorie van (begrijpend) lezen (vgl. Barrett, 1968). Bovendien worden de gepostuleerde vaardigheden voor met name begrijpend lezen vaak in vage gedragster-

men geformuleerd zonder dat de kenmerken van het object van het gedrag i.c. de tekst worden vermeld (vgl. Gray, 1960; Davis, 1968). Een ander probleem vormen de tests die niet voldoende worden gevalideerd en die wat betreft betrouwbaarheid vaak voor verbetering in aanmerking komen. Een volgend probleem heeft betrekking op het feit dat er tot dusverre weinig bekend is omtrent het effect van de verschillende programma-onderdelen (units) op de beheersing van leesvaardigheden en op begrijpend lezen in het algemeen. Een fundamenteel probleem van de "skill-approach" betreft de vraag of begrijpend lezen één algemene vaardigheid met niet te onderscheiden deelvaardigheden is óf een samenhangend geheel van vaardigheden. Indien het laatste het geval is, rijst de vraag welke relaties er bestaan tussen de onderscheiden vaardigheden en wat het effect is van het verwerven van vaardigheden op andere vaardigheden (vgl. Chapman, 1971). Het laatst genoemde probleem heeft rechtstreeks betrekking op de legitimiteit van de "skill-approach" voor begrijpend lezen. Tijdens de ontwikkeling van de voorlopige lijst van doelstellingen van het leesonderwijs werden we voortdurend geconfronteerd met de vraag naar de onderscheiding van vaardigheden (vgl. Aarnoutse et al., 1976).

Hoewel er meer problemen ten aanzien van deze benadering kunnen worden geformuleerd, zullen we ons in het onderhavige onderzoek beperken tot de bovengenoemde. Hierbij gaan we uit van de veronderstelling dat de "skill-approach" een positieve bijdrage kan leveren aan de planning en uitvoering van het begrijpend leesonderwijs in de lagere school. Omdat de term "skills" op verschillende zaken zoals doelen, activiteiten of gedragingen, of processen betrekking kan hebben, geven we in dit onderzoek de voorkeur aan de term "taken" en "taakgericht onderwijs". Taken worden in dit verband opgevat als opdrachten die de leerlingen ten aanzien van een tekst moeten leren uitvoeren. Onder taakgericht onderwijs op het gebied van begrijpend lezen verstaan we die vorm van leesonderwijs waarbij de leerkracht expliciet aandacht besteedt aan één bepaalde leestaak of een samenhangend geheel van leestaken en de leerlingen stapsgewijs leert hoe een dergelijke taak of een samenhangend geheel van taken kan worden uitgevoerd.

In aansluiting op de eerder vermelde problemen van de "skill-approach" kan nu de opzet en het doel van het onderzoek worden beschreven.

In de eerste fase van het onderzoek wordt een poging gedaan om drie inferentiële leestaken te selecteren op basis van een theorie van begrijpend lezen en inferentieprocessen. Een uitgebreide literatuurstudie is hiervoor

noodzakelijk. Na de selectie worden de taken zo goed mogelijk gespecificeerd en geoperationaliseerd in de vorm van drie taakgerichte tests. Omdat deze tests in de tweede fase van het onderzoek meerdere malen bij dezelfde proefpersonen (leerlingen van het vierde leerjaar) worden afgenomen, worden er parallelvormen ontwikkeld. Hetzelfde geldt voor twee typen cloze tests. Vervolgens worden voor elk van de drie geselecteerde taken experimentele programma's of leereenheden ontwikkeld. Elk programma is afzonderlijk gericht op het leren beheersen van een inferentiële leestaak, waarbij - zoals eerder vermeld - informatie uit een tekst moet worden afgeleid. De programma's omvatten ongeveer 10 lessen, die in afwijking van het Wisconsin Design klassikaal worden aangeboden. De programma's worden in enkele vierde klassen beproefd en waar nodig veranderd. Met behulp van de taakgerichte tests wordt een indicatie verkregen omtrent het effect van elk programma op de beheersing van de betreffende leestaak.

De tweede fase van het onderzoek bestaat uit twee onderscheiden delen nl. een descriptief gedeelte en een experimenteel gedeelte. Het descriptief gedeelte heeft vooral tot doel de betrouwbaarheid en validiteit van de vijf ontwikkelde paralleltests vast te stellen bij leerlingen van het vierde leerjaar. Voor de bepaling van de betrouwbaarheid van de tests worden de vier hoofdmethoden die Drenth (1972) onderscheidt, toegepast. In verband met het validiteitsonderzoek zal de samenhang worden vastgesteld tussen de vijf tests onderling en tussen deze tests en gestandaardiseerde tests voor technisch lezen, begrijpend lezen, woordenschat en non-verbale intelligentie. Bovendien zal de samenhang worden vastgesteld tussen de scores op de ontwikkelde tests en rapportcijfers voor lezen, taal en rekenen. Behalve de betrouwbaarheid en validiteit van de tests zullen ook de prestaties van de leerlingen op deze tests worden bepaald en de samenhang tussen deze prestaties met variabelen als sociaal milieu en school. Ook wordt met deze tests de verandering in prestaties over drie meetmomenten vastgesteld. In het experimentele gedeelte van het onderzoek zal worden nagegaan wat de waarde is van taakgericht onderwijs voor begrijpend lezen bij leerlingen van het vierde leerjaar. In een onderzoeksopzet, die getypeerd kan worden als een "true experimental pretest-posttest control group design", zal worden vastgesteld wat het effect is van elk taakgericht programma op de beheersing van de desbetreffende taak, gemeten met de daarmee corresponderende test. Ook zal worden nagegaan wat het effect is van elk programma op begrijpend lezen in het algemeen, zo-

als gemeten met de beide cloze tests. In het experimentele gedeelte zal bovendien worden nagegaan of de drie inferentiële leestaken in werkelijkheid zijn te onderscheiden. Vastgesteld zal worden wat het transfer-effect is van elk programma op de beheersing van die taken waarop de beide andere programma's zijn gericht en die gemeten worden met de daarmee corresponderende tests. Het antwoord op de laatste vraag is van essentieel belang voor taakgericht leesonderwijs. Indien zou blijken dat de taken in werkelijkheid niet zijn te onderscheiden, vervalt een belangrijk motief voor deze vorm van leesonderwijs.

Uit het bovenstaande zal duidelijk zijn dat positieve resultaten van het experimentele gedeelte van het onderzoek een belangrijk motief kunnen vormen voor een andere planning en uitvoering van het begrijpend leesonderwijs in de lagere school en wel in de richting van taakgericht onderwijs.

Het doel van het onderzoek kan als volgt worden geformuleerd:

- a. de prestaties en de verandering in prestaties vaststellen van leerlingen van het vierde leerjaar op vijf ontwikkelde tests voor het begrijpend lezen van teksten;
- b. de betrouwbaarheid en validiteit vaststellen van vijf ontwikkelde tests voor het begrijpend lezen van teksten;
- c. het effect vaststellen van drie ontwikkelde experimentele programma's (leereenheden), die elk afzonderlijk zijn gericht op een betere beheersing van één bepaalde taak op het gebied van het begrijpend lezen van teksten;
- d. de realiteitswaarde vaststellen van de onderscheiding van drie taken op het gebied van het begrijpend lezen van teksten.

Opgemerkt moet worden dat het doel genoemd onder (a) van secundair belang is. In hoofdstuk IV en V wordt het doel van het onderzoek nader geconcretiseerd in de vorm van vraagstellingen en hypothesen.

1.3. Overzicht van het rapport

Dit rapport bevat in totaal zes hoofdstukken. In *hoofdstuk I* wordt een beschrijving gegeven van de voorafgaande onderzoeken en van de aanleiding en het doel van het onderhavige onderzoek.

Hoofdstuk II bevat onder meer het resultaat van de literatuurstudie die in verband met de selectie van de leestaken werd verricht. Na een inleiding

over enkele typen leesmodellen, worden twee benaderingen van het begrijpend leesproces aan de orde gesteld evenals de problematiek van inferenties en inferentieprocessen. Hierna volgt een beschrijving van de selectie en specificatie van de leestaken.

Hoofdstuk III bevat het resultaat van de literatuurstudie die in verband met de drie belangrijkste onderzoeksvragen werd uitgevoerd. Eerst wordt aandacht besteed aan programma's (curricula) van begrijpend lezen en aan onderzoek naar het effect van strategieën en experimentele programma's. Daarna volgt een beschrijving van de vaardigheden die een aantal auteurs aan begrijpend lezen onderscheiden en een overzicht van onderzoek met betrekking tot vaardigheden op het gebied van begrijpend lezen. Tenslotte geven we een beschrijving van de verschillende methoden voor het meten van begrijpend lezen en besteden we aandacht aan de twee typen tests die hier centraal staan en aan onderzoek dat in verband met deze tests van belang is.

Hoofdstuk IV heeft betrekking op de opzet en de methode van het onderzoek. Na de formulering van de vraagstellingen van het descriptief en experimenteel gedeelte van het onderzoek volgt de onderzoeksopzet en wordt een beschrijving gegeven van de ontwikkeling van de tests en de programma's. Daarna komt de selectie van de scholen aan de orde evenals een vergelijking van de experimentele scholen en controlescholen wat betreft het leesonderwijs. Tenslotte wordt een beschrijving gegeven van de dataverzameling en van de wijze waarop het experiment is voorbereid en uitgevoerd.

Hoofdstuk V heeft betrekking op de analyse van de gegevens en de resultaten van het onderzoek. In het eerste gedeelte van dit hoofdstuk worden de vraagstellingen c.q. verwachtingen van het descriptieve gedeelte van het onderzoek nader gepreciseerd en successievelijk beantwoord of getoetst. In het tweede deel van dit hoofdstuk worden eerst de hypothesen van het experimentele gedeelte van het onderzoek geformuleerd. Daarna volgt een beschrijving van de analyse van de data. Tenslotte worden de belangrijkste conclusies geformuleerd.

In *Hoofdstuk VI* worden eerst de belangrijkste resultaten van het onderzoek in het kort vermeld en waar nodig van commentaar voorzien. Daarna volgen enkele beperkingen van het onderzoek. Tot slot worden suggesties gegeven voor verder onderzoek.

Dit hoofdstuk bevat onder meer het resultaat van een literatuurstudie die in het begin van het onderhavige onderzoek is uitgevoerd met het doel inferentiële leestaken te selecteren op basis van een theorie van begrijpend lezen en van inferentieprocessen. Na een inleiding die handelt over enkele typen leesmodellen, worden in par. 2.2. twee belangrijke benaderingen van het proces van begrijpend lezen aan de orde gesteld. Bij elke benadering wordt nagegaan in hoeverre deze als basis kan dienen voor de selectie van de leestaken. Daarna volgt in par. 2.3. een beknopt overzicht van de problematiek van inferenties en inferentieprocessen met als conclusie dat een algemene theorie over inferentieprocessen vooralsnog ontbreekt. In par. 2.4. wordt een beschrijving gegeven van de selectie en specificatie van de leestaken. Deze paragraaf bevat tevens de belangrijkste resultaten van de literatuurstudie die in verband met de specificatie van de taken werd uitgevoerd.

2.1. Inleiding

De laatste tien jaar wordt in de literatuur over lezen vrij veel aandacht besteed aan theorieën en modellen van het leesproces. In 1970 verscheen de eerste uitgave van "Theoretical models and processes of reading" onder redactie van Singer en Ruddell. Eén jaar later verscheen een uitvoerig rapport met als titel "The literature of research in reading with emphasis on models" (Davis, 1971). Dit rapport was een van de laatste publicaties van het groots opgezette en niet voltooide "Targeted research and development program in reading". In 1972 presenteerde Geyer op basis van dit project een overzicht van 48 modellen van het leesproces en daarmee verwante processen, die volgens hem van belang konden zijn voor verder onderzoek. In 1975 besteedden Gibson en Levin de nodige, kritische aandacht aan twee typen leesmodellen nl. informatietheoretische en psycholinguïstische model-

len. Uitgaande van het idee dat er veel soorten leesprocessen zijn, verwerpen zij de mogelijkheid om één model voor hét leesproces te ontwikkelen. In een overzichtsartikel van 1978 besteedden Calfee en Drum - gedeeltelijk als reactie op de publicatie van Gibson en Levin (1975) - veel aandacht aan de problematiek van modellen van het leesproces en van het leesleerproces. Volgens hen kunnen de huidige leesmodellen en -theorieën in drie categorieën worden verdeeld nl. bottom-up modellen, top-down modellen en combinaties van beide benaderingen.

Een prototype van een bottom-up model heeft Gough (1972) ontwikkeld. In dit model wordt het leesproces opgevat als een geordende reeks van transformaties die strikt sequentiëel verlopen. Stap voor stap wordt overgegaan van het ene niveau van analyse naar het volgende, hogere niveau. Een lager niveau van analyse kan wel invloed hebben op een direct daarop volgend niveau, maar omgekeerd kunnen hogere niveaus geen invloed uitoefenen op de analyse van lagere niveaus. Volgens Gough verloopt het leesproces van letter naar letter, van woord naar woord en van woordgroep naar woordgroep. Dit serieel model van lezen, dat geen interactie toestaat van hogere-orde processen met lagere-orde processen, is door verschillende onderzoekers als Brewer (1972), Gibson en Levin (1975), McConkie en Rayner (1976) en Calfee en Drum (1978) bekritiseerd. Rumelhart (1977a) verwerpt het model van Gough evenals het model van LaBerge en Samuels (1974), dat ook als een bottom-up model kan worden opgevat. Volgens Rumelhart kan een bottom-up model geen voldoende verklaring geven voor het feit dat de perceptie van letters vaak wordt beïnvloed door de context waarin de letters verschijnen. Evenmin kan een dergelijk model een voldoende verklaring geven voor het feit, dat de perceptie van woorden vaak wordt beïnvloed door de semantische context en dat de betekenis van een woord of van een zin vaak afhankelijk is van de andere woorden of van de andere zinnen.

Goodman (1976) typeert de opvatting van Gough als "the end of the nose view" (pag. 503). Volgens hem is lezen een psycholinguistisch proces, een wisselwerking tussen gedachte en taal. Het is een proces van voorspellen, selecteren en toetsen, met andere woorden een "psycholinguistic guessing game". Op basis van voorafgaande informatie en reeds verworven kennis voorspelt de geoefende lezer wat zal volgen en selecteert hij juist voldoende informatie om zijn voorspelling te toetsen. Lezen is volgens Goodman (1970;

1973; 1976) een reconstructieproces waarbij de geoefende lezer verwachtingen ontwikkelt en deze toetst door minimaal gebruik te maken van grafische aanwijzingen. Een dergelijk model van lezen, dat de kennis en de verwachtingen van de lezer centraal stelt, is een prototype van een top-down model. Gibson en Levin (1975), McConkie en Rayner (1976) en Calfee en Drum (1978) wijzen een dergelijk model van de hand. Volgens Gibson en Levin (1975) is het model van Goodman niet specifiek genoeg en geeft het op een aantal belangrijke vragen geen duidelijk antwoord. Volgens McConkie en Rayner (1976) wordt in dit model niet duidelijk gemaakt waarover de lezer voorspellingen doet en hoe hij deze toetst. De veronderstelling als zou het Engels proza voldoende redundant zijn om het volgende woord te voorspellen, wijzen zij van de hand, evenals de veronderstelling dat het eenvoudiger is om een hypothese te ontwikkelen over de grafische input dan gewoon te kijken naar wat er staat (vgl. Reitsma, 1979).

Terwijl een bottom-up model een sterk accent legt op de analyse van visuele informatie, gaat een top-down model uit van de veronderstelling dat hogere cognitieve processen het leesproces en dus ook de analyse van de visuele informatie in sterke mate beïnvloeden en sturen (vgl. Stanovich, 1980). Een derde categorie van leesmodellen gaat noch van een strikte bottom-up benadering, noch van een strikte top-down benadering uit. Een goed voorbeeld uit deze categorie, dat niet als een integratie van de beide andere benaderingen mag worden opgevat, is het model van Rumelhart (1977a). Volgens Rumelhart is het leesproces een perceptueel-cognitief proces, waarbij tegelijkertijd gebruik wordt gemaakt van sensorische, orthografische, lexicale, syntactische en semantische informatie. Een dergelijk interactiemodel verschilt niet alleen van een bottom-up model maar ook van een top-down model. In een top-down model sturen de hogere-orde processen (de processen van begrijpen) de lagere orde-processen (de decodeerprocessen), terwijl in een interactiemodel de hogere orde-processen de alternatieven van de lagere orde-processen inperken, maar ook zelf worden ingeperkt door de lagere orde-processen. In de volgende paragraaf worden twee benaderingen van begrijpend lezen aan de orde gesteld, die allebei tot de derde categorie moeten worden gerekend.

2.2. Twee benaderingen van begrijpend lezen

Uit de voorgaande paragraaf blijkt dat de twee contrasterende modellen verschillende aspecten van het leesproces benadrukken. Terwijl het model van Gough een sterk accent legt op het decodeerproces i.c. woordherkenning, staat in het model van Goodman het begrijpen (comprehension) van de tekstuele informatie centraal. Zoals Gibson en Levin (1975) opmerken, is er veel meer bekend over processen die een rol spelen bij woordherkenning dan over processen die van invloed zijn op het begrijpen van zinnen en teksten. Twee redenen worden hiervoor aangevoerd, nl. de moeilijkheid om de eenheden te bepalen waarin de semantische informatie verdeeld kan worden én de geringe progressie van een theorie over begrijpen. Toch is er de laatste tien jaar vooruitgang geboekt op het gebied van begrijpend lezen of ruimer opgevat tekstverwerking (vgl. Kintsch, 1974; Winograd, 1977). Vanuit de cognitieve psychologie, de linguïstiek en het artificieel intelligentie-onderzoek zijn verschillende min of meer uitgewerkte voorstellen of modellen ontwikkeld en is een grote verscheidenheid van onderzoek verricht. Hoewel de theorievorming nog in een beginfase verkeert en de experimentatie vaak nog voor allerlei uitleg vatbaar is, kan men in de zich snel uitbreidende literatuur enkele lijnen ontdekken. In het volgende zullen we twee verschillende benaderingen van het begrijpend leesproces in het kort behandelen en nagaan in hoeverre deze benaderingen als basis kunnen dienen voor de selectie van de inferentiële leestaken.

Alvorens de beide benaderingen aan de orde te stellen, besteden we eerst enige aandacht aan het theoretisch begrip "begrijpend lezen", zoals dat door verschillende auteurs wordt gedefinieerd. De term "begrijpend lezen" is een pleonasme voor degene die het begrijpen als de kern van het leesproces beschouwt. Omdat in het verleden het leesproces vaak werd beperkt tot de reconstructie van de klankvorm van een woord op basis van zijn grafische weergave (vgl. Elkonin, 1963; Harris, 1969), is de term "reading comprehension" of begrijpend lezen in de Amerikaanse en Nederlandse literatuur geïntroduceerd en gangbaar geworden.

In een artikel van 1917 getiteld "Reading as reasoning: a study of mistakes in paragraph reading" omschrijft Thorndike het begrijpen van een korte tekst als een denkproces, waarbij informatie wordt geselecteerd, af-

geleid en met elkaar gerelateerd. "Understanding a paragraph is like solving a problem in mathematics. It consists in selecting the right elements of the situation and putting them together in the right relations, and also with the right amount of weight or influence or force for each" (1971, pag. 431). Volgens Thorndike moet het onderwijs veel meer aandacht besteden aan het begrijpend lezen in de vorm van stillezen, dan aan het technisch hardop lezen. Deze gedachte vinden we eerder bij Huey (1908) en later in ons land bij Van Veen (1929) en Van der Hoeve, Kohnstamm en Van Veen (1933).

Volgens Carroll (1972) kan het begrijpen van gesproken en geschreven taal worden beschouwd als een proces dat uit minstens twee fasen bestaat nl. het begrijpen van de linguïstische informatie en het relateren van deze informatie aan een bredere context. Carroll pleit er voor beide fasen zo goed mogelijk van elkaar te onderscheiden en te onderzoeken.

Trabasso (1972) is van mening dat het begrijpen van gesproken en geschreven taal een reeks van psychische processen omvat. Deze processen bestaan uit een serie mentale operaties die de linguïstische informatie verwerken. Volgens hem zijn bij het begrijpen van zinnen twee operaties van belang nl. de encoding van de informatie in interne representaties en de vergelijking van deze representaties (bijvoorbeeld de representatie van een zin met een afbeelding).

Bransford en McCarrell (1974) karakteriseren het begrijpen van gesproken en geschreven taal als "the grasping of relations" (pag. 200). Volgens deze onderzoekers is het begrijpen van talige informatie afhankelijk van de cognitieve bekwaamheid van de luisteraar of lezer om kennis te activeren die hem of haar in staat stelt de juiste verbanden te leggen (vgl. Harris, 1979).

Clark (1977) beschouwt het begrijpen van gesproken of geschreven taal als een probleemoplossend proces, waarbij de luisteraar of lezer één of meer strategieën toepast om de boodschap van de spreker of schrijver te relateren aan de informatie waarover hij beschikt (vgl. Olshavsky, 1976-1977).

In Deelrapport 3 (Aarnoutse et al., 1977) hebben wij het begrijpen gekarakteriseerd als een proces dat relationeel van aard is en noodzakelijkerwijs gebonden aan het referentie- en interpretatiekader van de luisteraar of lezer. Begrijpend lezen omschrijven we als een proces waarbij de lezer de nieuwe informatie relateert aan de bestaande kennisstructuur.

Bol (1978) typeert het begrijpend lezen als een creatieve bezigheid met een sterk probleemoplossend karakter. "Begrijpend lezen kan worden gezien als een complexe vaardigheid berustend op een systeem van door oefening onderling gecoördineerde deelvaardigheden, die een rol spelen bij het gebruiken en toepassen van kennis in het kader van leestaken" (pag. 203).

Van Oostendorp en Den Uyl (1980) omschrijven het begrijpen van een (gelezen) tekst als het construeren van een interne representatie van een tekst. Volgens deze onderzoekers moet de representatie van een tekst voldoende hecht en rijk zijn, wil er sprake zijn van een juist begrip van de tekst.

Uit de bovenstaande opsomming van een aantal definities blijkt dat het begrijpend leesproces in vrij globale termen wordt omschreven en dat het merendeel van de auteurs begrijpend lezen opvat als een actief proces waarbij de juiste relaties moeten worden opgespoord.

We zullen nu twee benaderingen van het begrijpend leesproces in het kort aan de orde stellen nl. de "text based" benadering van Kintsch en Van Dijk (1978) en de schemabenadering van onder anderen Rumelhart en Norman (1978) en Anderson (1978). De eerste benadering vindt haar vertrekpunt in de tekst of de semantische structuur van de tekst en veronderstelt dat het begrijpend leesproces primair gestuurd wordt vanuit de tekst (vgl. Winograd, 1977). De tweede benadering richt zich primair op de cognitieve processen die bij de tekstverwerking een rol spelen en veronderstelt dat de aanwezige kennis-structuur met name schemata hierbij een bepalende factor is. Opgemerkt moet worden dat beide benaderingen elkaar niet uitsluiten, maar wellicht aanvullen (vgl. Kok, Boonman & Beukhof, 1980).

2.2.1. Het model van Kintsch en Van Dijk

Volgens Kintsch en Van Dijk (1978) kan de semantische structuur van een tekst op twee niveaus worden beschreven nl. op het niveau van de microstructuur en op het niveau van de macrostructuur. De microstructuur vormt de structuur van de afzonderlijke proposities en hun relaties. De macrostructuur is meer globaal van aard. In het volgende zullen we eerst aandacht besteden aan deze beide structuren evenals aan de schematische structuren van teksten. Daarna komt het model van "text comprehension" aan de orde.

Volgens Kintsch (1974; 1977a en 1977b) kan de betekenis van een tekst op het microniveau worden gerepresenteerd door middel van een tekstbasis (text base). Een tekstbasis is een geordende lijst of sequentie van proposities, die zijn samengesteld uit concepten. Elke propositie bevat een relationeel concept of predicaat en een of meer argumenten, die bestaan uit concepten of andere proposities. Een relationeel concept specificceert een relatie tussen de argumenten en wordt in een tekst meestal gerealiseerd door werkwoorden, bijvoeglijke naamwoorden, bijwoorden en verbindingswoorden. De argumenten van een propositie vervullen verschillende semantische functies zoals die van agens, object en doel. Een tekstbasis is een coherente, gestructureerde eenheid. Een van de criteria voor de semantische coherentie van een tekstbasis (en van een tekst) is volgens Kintsch en Van Dijk (1978) de referentiële coherentie. Deze lineaire vorm van coherentie komt overeen met de herhaling of overlap van argumenten tussen proposities. Twee proposities van een tekstbasis kunnen met elkaar worden verbonden doordat in beide een zelfde argument voorkomt. Hoewel de referentiële coherentie een belangrijk criterium is voor de coherentie van een tekstbasis is zij linguïstisch gezien niet noodzakelijk noch voldoende. De proposities van een tekstbasis kunnen niet alleen referentiëel met elkaar zijn verbonden, maar ook intentioneel of conceptueel. Een tekstbasis kan nl. ook coherent zijn omdat elke propositie is gerelateerd aan het onderwerp (topic) of thema van een tekst (Van Dijk, 1977). Met deze notie van het onderwerp van een tekst zijn we gekomen aan het tweede niveau nl. de macrostructuur van een tekst.

De macrostructuur is een representatie van de betekenis van een tekst op een meer globaal niveau. Evenals de microstructuur wordt de macrostructuur van een tekst beschreven in termen van proposities en sequenties van proposities en vormt zij een coherent, gestructureerd geheel. De macrostructuur wordt afgeleid van de tekstbasis of microstructuur en representeert met de macroproposities de kern van de tekstbasis. De betekenis van de tekst als geheel blijft hierbij echter intact. De relatie van de macrostructuur tot de microstructuur heeft betrekking op noties als belangrĳheid of relevantie; "the macro-interpretation defines the most important or essential object or event denoted by a sequence of propositions" (Van Dijk, 1977, pag. 8).

Van Dijk (1977; 1978) heeft een aantal semantische regels ontwikkeld om de proposities van een tekstbasis om te zetten in een reeks macroproposities. Deze zogenaamde macroregels reduceren en organiseren de meer gedetailleerde informatie van de microstructuur van een tekst en kunnen eventueel ook op sequenties van macroproposities worden uitgevoerd. (Dit laatste betekent dat er verschillende niveaus van macrostructuren mogelijk zijn). De belangrijkste beperking van de macroregels bestaat hierin dat geen propositie mag worden weggelaten die conditioneel is voor de interpretatie van een volgende propositie. Van Dijk (1978) onderscheidt de volgende macroregels:

a. Weglating.

Deze regel houdt in dat elke propositie die noch direct noch indirect van belang is voor de interpretatie van een volgende propositie kan worden weggelaten.

b. Generalisatie.

Deze regel betekent dat iedere sequentie van proposities, waarin concepten voorkomen die door een gemeenschappelijk superconcept worden omvat, door een propositie met dit superconcept wordt vervangen.

c. Constructie.

Deze regel houdt in dat iedere sequentie van proposities die normale voorwaarden, componenten of gevolgen van een globaal feit aanduiden, wordt vervangen door een propositie die dit globale feit aanduidt.

De vorming van een coherente macrostructuur van een tekst veronderstelt evenals de realisering van een coherente microstructuur, de nodige algemene en contextuele kennis.

Naast de bovengenoemde semantische structuren van teksten onderscheiden Kintsch en Van Dijk (1978) ook nog schematische structuren. Met deze term duiden zij de conventionele globale structuur aan die aan bepaalde teksttypen ten grondslag ligt. Voorbeelden van schematische structuren zijn de structuur van een verhaal (narrative), de structuur van een argumentatie en de structuur van een empirisch onderzoeksverslag. De schematische structuur (of superstructuur) van een verhaal bestaat volgens Van Dijk (1978) uit categorieën als plot, episode, setting, belevenis, complicatie en oplossing. Uit onderzoek van Kintsch (1977a) en Rumelhart (1977b) blijkt dat kennis van schematische structuren een factor is die bij de verwerking en productie van tekstuele informatie een rol speelt. Een lezer die een ver-

haal leest, verwacht bijvoorbeeld de episodische structuur die kenmerkend is voor verhalen. Bovendien kunnen schematische structuren een belangrijke controle-functie vervullen bij het onderzoek naar het proces van begrijpend lezen.

In hun model voor "text comprehension" gaan Kintsch en Van Dijk (1978) van de veronderstelling uit, dat het begrijpend lezen van teksten een constructief proces is waarbij een coherent geheel van proposities en macro-proposities wordt gevormd. Volgens hen verloopt het begrijpend leesproces van teksten in cycli. Tijdens één cyclus worden de proposities van één zin of woordgroep verwerkt. Gedurende zo'n cyclus worden bepaalde proposities geselecteerd, in de buffer van het korte termijn geheugen vastgehouden om vervolgens met bepaalde proposities van de volgende cyclus te worden verbonden. Als er in het korte termijn geheugen een verbinding wordt aangebracht tussen een van de nieuwe proposities en die welke in de buffer worden vastgehouden, met andere woorden als er een overlap van argumenten bestaat tussen de inputproposities en de proposities in de korte termijn buffer, dan wordt de input geaccepteerd als coherent met de voorgaande informatie en verloopt het leesproces vrijwel automatisch. Wanneer er echter geen verbinding kan worden aangebracht, dan moet een zoekproces naar alle reeds verwerkte proposities worden gestart. "In reading comprehension, the search includes all previous propositions because even those propositions not available from long-term memory can be located by re-reading the text itself" (Kintsch & Van Dijk, 1978, pag. 368). Pas wanneer dit zoeken geen succes heeft, wordt een inferentieproces geïnitieerd waarbij één of meer proposities aan de tekstbasis worden toegevoegd om een verbinding tussen de inputproposities en de reeds verwerkte proposities tot stand te brengen. Volgens Kintsch en Van Dijk (1978) worden bij het zoek- en inferentieproces de "resources" van de lezer relatief zwaar belast.

Een belangrijke vraag is uiteraard welke proposities tijdens een cyclus worden verbonden. Kintsch en Van Dijk (1978) beantwoorden deze vraag door er op te wijzen dat bij een goede selectiestrategie die proposities worden gekozen die belangrijk zijn en recent verwerkt. Een propositie die reeds met vele andere proposities is verbonden, is volgens hen waarschijnlijk meer relevant dan een propositie die tot dan toe geen belangrijke rol heeft gespeeld. Dit betekent in concreto dat proposities die behoren tot de ho-

gere niveaus van de tekstbasishiërarchie bij voorkeur voor selectie in aanmerking komen. Deze proposities worden ook beter onthouden dan proposities van een lager niveau. Volgens Kintsch en Van Dijk (1978) is dat te verklaren door het frequenter verwerken - via de opname in de buffer van het korte termijn geheugen - van de "hogere" proposities, waardoor de reproductie-waarschijnlijkheid toeneemt.

Het model zoals tot nu toe beschreven, heeft betrekking op de vorming van een coherent geheel van microproposities. Dit netwerk van proposities kan als een "graph" worden weergegeven. Van elke cyclus en van alle cycli samen kan een "graph" worden geconstrueerd. Binnen deze zogenaamde "coherence graphs" kunnen verschillende niveaus van proposities worden onderscheiden.

Tegelijk met de vorming van een coherente tekstbasis of microstructuur vinden er processen plaats waarbij macroproposities worden gevormd. De buffer van het korte termijn geheugen moet deze macroproposities vasthouden zodat er een coherente macrostructuur van de gehele tekst kan ontstaan.

De zogenaamde macro-operatoren zetten de proposities van een tekstbasis om in een reeks macroproposities die de kern van een tekst representeren. "They do so by deleting or generalizing all propositions that are either irrelevant or redundant and by constructing new inferred propositions" (Kintsch & Van Dijk, 1978, pag. 372). De macro-operatoren werken probabilistisch en cyclisch en worden gestuurd door de doelen van de lezer. Deze doelen bepalen uiteindelijk welke microproposities, generalisaties of constructies van microproposities relevant zijn en worden opgenomen in de macrostructuur van een tekst. In het model worden deze leesdoelen aangeduid met de term "schema".

Volgens Kintsch en Van Dijk (1978) zijn de macroprocessen alleen voorspelbaar wanneer het schema expliciet kan worden gemaakt. Twee situaties voldoen volgens hen aan deze eis nl. de situatie waarbij een tekst met een duidelijke schematische structuur wordt aangeboden en de situatie waarbij een lezer een duidelijk doel voor ogen heeft. In beide situaties werkt het schema top-down. Dit betekent overigens niet dat het model een top-down model is (Kintsch, 1977a). Zoals we gezien hebben wordt de lineaire samenhang via de invoer van een sequentie van proposities bottom-up tot stand gebracht. De macrostructuur wordt deels bepaald door de bottom-up verkre-

gen referentiële coherentie en deels door het top-down werkende schema.

Het model van Kintsch en Van Dijk (1978) heeft niet alleen betrekking op het begrijpend lezen van teksten maar ook op de productie van teksten. Volgens de ontwerpers is dit productieproces gedeeltelijk reproductief en gedeeltelijk constructief. In een herinneringsprotocol of een samenvatting kan men onder andere proposities uit de micro- en macrostructuur verwachten (reproducties), proposities die zijn afgeleid uit andere proposities (reconstructies) en proposities over bijvoorbeeld de inhoud van de tekst, inclusief commentaar (meta-uitspraken).

Het testen van het bovengeschetste model is mogelijk door de herinneringsfrequenties van de verschillende proposities te bepalen. Daarbij kan het best passende model gevonden worden door de drie parameters n (maximum input per cyclus), s (capaciteit van de korte termijn buffer) en p (reproductiewaarschijnlijkheid) te variëren. Met name is n afhankelijk van de kennis die de lezer heeft over het onderwerp van de tekst. Ook s en p zijn gedeeltelijk afhankelijk van de bekendheid van de lezer met het onderwerp.

Het boven beschreven model gaat uit van enkele veronderstellingen en is gedeeltelijk gebaseerd op resultaten van empirisch onderzoek. Het veronderstelt dat er een representatie van de tekst in het geheugen wordt opgeslagen die episodisch van aard is, d.w.z. aan de context van het lezen van die tekst gebonden. Bovendien veronderstelt het model dat de verbindingen tussen proposities in het korte termijn geheugen tot stand komen en dat dit geheugen slechts een beperkt aantal proposities tegelijk kan verwerken (vgl. Van Oostendorp & Den Uyl, 1980). Wat betreft de empirische fundering kan onder andere worden gewezen op het onderzoek van Kintsch et al. (1975) dat aanwijzingen bevat dat referentiële coherentie of overlap van argumenten een rol speelt bij de vorming van een coherente tekstbasis. Verder kan gewezen worden op de experimenten van Kintsch (1974) waarin deze probeert aan te tonen dat het proces van begrijpen evenals het herinneringsproces een actief proces is waarbij informatie wordt geselecteerd en geordend (vgl. Thomassen & Kempen, 1976). Uit onderzoek van Kintsch (1977a) blijkt dat kennis van de schematische structuur van een tekst een rol speelt bij de verwerking van informatie. Wat betreft de vraag of bij het lezen van langere teksten inderdaad sprake is van macroprocessen kan worden verwezen naar onderzoek van onder anderen Kintsch en Van Dijk (1978).

Het zou te ver voeren om de bovengenoemde onderzoeken hier te bespreken. Een uitzondering willen we maken voor het laatstgenoemde onderzoek, waarin het model wordt toegepast op een paragraaf van een psychologisch onderzoeksrapport. In dit experiment kregen volwassen proefpersonen het betreffende rapport van 1300 woorden te lezen met daarna de opdracht hiervan een herinneringsprotocol en een samenvatting te maken. Andere proefpersonen kregen deze opdracht na één maand en drie maanden. Uit de analyse van de data van de eerste paragraaf bleek dat bij onmiddellijke herinnering het aantal reproducties drie maal zo groot was als het aantal reconstructies, terwijl na drie maanden bijna evenveel reproducties als reconstructies voorkwamen. Bij onmiddellijke herinnering was het aantal macroproposities vier maal zo groot als het aantal microproposities; deze ratio nam met de tijd toe zodat na drie maanden de herinneringsprotocollen in dit opzicht sterk op samenvattingen leken. Wat betreft de samenvattingen kwam ongeveer hetzelfde beeld naar voren, hoewel minder geprononceerd. Het percentage reproducties daalde over de tijd gezien slechts weinig (van 72% naar 61%). De meeste reproducties bestonden, zoals voorspeld, uit macroproposities; de waarschijnlijkheid dat een macropropositie in de samenvatting voorkwam was 12 maal groter dan een micropropositie. Uit dit onderzoek blijkt dat bij het lezen van langere teksten macroprocessen plaatsvinden. Kintsch en Van Dijk (1978) noemen de gevonden data bemoedigend en voegen er aan toe dat verder onderzoek en een verdere uitbreiding van hun model noodzakelijk is.

Het model kent een aantal beperkingen die door de ontwerpers zelf gedeeltelijk ook worden onderkend. Zo gaat het model uit van semantische representaties zonder dat een "parser" wordt aangegeven die de zinnen van een tekst omzet in semantische representaties. Bovendien geeft het model niet aan hoe inferenties op basis van algemene kennis plaatsvinden. Het model geeft wel aan dát inferenties plaatsvinden en welke beperkte rol ze spelen, maar niet hoe het inferentieproces verloopt. Een andere beperking is dat het model alleen toepasbaar is op die situatie waarbij sprake is van een schema. Dit betekent vaak een beperking tot teksten met een duidelijke schematische structuur. Verder wordt niet duidelijk gemaakt hoe het begrijpend leesproces door het schema wordt gereguleerd, met andere woorden hoe het toetsingsmechanisme in feite werkt. Zoals we reeds opmerkten, geeft het model van Kintsch en Van Dijk weinig informatie omtrent inferenties en inferentiepro-

cessen. Dit betekent in feite dat het model voor het onderhavige onderzoek niet als basis kan dienen voor de selectie van inferentiële leestaken. Uit het model kunnen wel enkele inferentiële taken worden afgeleid die onder andere een rol spelen bij de vorming van de macrostructuur van een tekst. Zo kan bij de vorming van constructies en reconstructies en bij het maken van generalisaties sprake zijn van het afleiden van informatie. Uit het vervolg zal blijken dat een van de geselecteerde taken betrekking heeft op het maken van generalisaties die niet in de tekst voorkomen.

2.2.2. De schematheorie van Rumelhart e.a.

Zoals reeds eerder vermeld, is de schematheorie van onder anderen Rumelhart en Norman (1978) en Anderson (1978) meer gericht op de cognitieve processen die tijdens de productie en verwerking van een tekst (discourse) plaatsvinden en minder op de semantische structuur van een tekst. Verondersteld wordt dat de kennis waarover iemand beschikt een belangrijke factor is bij het begrijpen en onthouden van teksten. Volgens deze theorie worden bij het lezen van een tekst bepaalde georganiseerde kenniseenheden in het geheugen geactiveerd en wordt de informatie van een tekst verwerkt door deze te koppelen aan de betreffende kenniseenheden. In het volgende zullen we dit idee uitwerken door eerst aandacht te besteden aan het begrip "schema" en de rol die schema's of schemata vervullen bij het begrijpen van teksten. Daarna zullen we een aantal onderzoeken bespreken. Tot slot volgen enkele kritische opmerkingen en een antwoord op de vraag in hoeverre de schematheorie als basis kan dienen voor de selectie van inferentiële leestaken.

Volgens Rumelhart en Norman (1978) bevat het menselijk geheugen twee soorten kennis nl. kennis die gebonden is aan een bepaalde specifieke situatie (bijvoorbeeld de herinnering van een etentje in een restaurant) en kennis die meer algemeen is waarbij geabstraheerd is van de kennis van specifieke situaties naar een klasse van situaties (bijvoorbeeld kennis omtrent eten in een restaurant). Deze algemene kennis, die van toepassing is op een grote klasse van situaties, is in het geheugen opgeslagen in de vorm van georganiseerde kenniseenheden of schemata. Schemata zijn volgens Rumelhart en Norman (1978) actieve, geïnterrelateerde kennisstructuren, die betrokken zijn bij het begrijpen van inkomende informatie en die de uitvoering van verwerkingsprocessen sturen.

Rumelhart en Ortony (1977) noemen een aantal kenmerken van schemata die volgens hen essentieel zijn. Een van de belangrijkste kenmerken is dat schemata generieke concepten representeren, die variëren in niveaus van abstractie. Een schema geeft generieke, algemene kennis weer, dat wil zeggen het representeert wat algemeen als waar wordt aangenomen voor objecten, situaties, gebeurtenissen, sequenties van gebeurtenissen, acties en sequenties van acties. Rumelhart en Ortony (1977) geven voorbeelden van eenvoudige, minder complexe schemata zoals het schema voor "geven" en "breken" en voorbeelden van meer complexe schemata zoals het schema voor "gezicht" en "oplossen van problemen". Volgens Rumelhart en Ortony (1977) bestaat het menselijk geheugen uit talloze informatiepakketten of schemata, die naar elkaar verwijzen. Zulke pakketten of schemata representeren volgens hen kennis op alle niveaus van abstractie "ranging from basic perceptual elements, such as the configuration of lines which form a square, to abstract conceptual levels which allow us to give cogent summaries of sequences of events occurring over substantial periods of time" (pag. 110).

Een ander kenmerk van schemata is al vermeld nl. dat schemata kennis representeren die geassocieerd is met concepten en niet met woorden. Rumelhart en Ortony (1977) benadrukken dat schemata geen linguïstische entiteiten zijn maar abstracte, symbolische representaties van kennis.

Een volgend kenmerk van schemata bestaat hierin dat ze elkaar kunnen insluiten. Een schema is een abstracte kennisstructuur, die bestaat uit een aantal componenten, elementen of constituenten, die zelf weer schemata zijn. De schemata die ingebed zijn in een dominant schema worden subschemata genoemd. Zo bevat het gezicht-schema een aantal subschemata als ogen, oren, mond en dergelijke. De inhoud van zo'n schema wordt gevormd door de specificatie van de normale constituerende delen, de subschemata, en de specificatie van de relaties die normaal tussen deze delen bestaan. Oog is een subschema van het gezicht-schema, terwijl pupil een subschema is van het oog-schema. Op deze wijze kan een hele hiërarchie van schemata en subschemata worden opgebouwd. Een voordeel van "embedded" schemata is dat een situatie of object begrepen kan worden in termen van belangrijke constituenten zonder noodzakelijke verwijzing naar de interne structuur van de constituenten zelf. Van de andere kant is de mogelijkheid aanwezig dat in een bepaalde situatie de interne structuur van een subschema kan worden geactiveerd.

Een vierde kenmerk van schemata is dat ze variabelen bevatten. Variabelen verwijzen naar algemene klassen van concepten die bij het bepalen van de implicaties van een schema voor een bepaalde situatie of tekst in de plaats kunnen komen voor de variabelen. Specifieke informatie wordt in het geheugen geëncodeerd wanneer constanten - specifieke waarden of specifieke concepten - in de plaats komen van de variabelen van een schema. Rumelhart en Norman (1978) illustreren het bovenstaande aan de hand van het schema "farming". Dit schema bevat variabelen als land, personen, gewas, dieren, machines en dergelijke. Als we nu lezen dat de Janssens een veeteeltbedrijf hebben, dan vervangen we de variabele "personen" voor de Janssens als de groep die de rol van boeren in het schema vervullen en de variabele "dieren" voor koeien, varkens en dergelijke. In dit voorbeeld worden slechts enkele variabelen van het schema "farming" vervangen of "geïncantiëerd" door constanten. Andere variabelen blijven ongespecificeerd. Op grond van onze algemene kennis van een veeteeltbedrijf kunnen we de andere variabelen geheel of ten dele invullen. Deze invulling is niet geheel vrij, maar gebonden aan de restricties van het betreffende schema. In navolging van Minsky (1975) spreken Rumelhart en Norman (1978) van de "default" waarden die bepaalde variabelen in een schema kunnen aannemen. Dit zijn waarden die aan bepaalde variabelen kunnen worden toegekend tenzij anders vermeld. Variabelen en hun restricties hebben de volgende functies: (a) ze specificeren de omvang van de objecten die de posities van de verschillende variabelen kunnen vullen, (b) ze maken het mogelijk om juiste gissingen over mogelijke waarden te maken, wanneer specifieke informatie over de variabelen ontbreekt.

De bovenstaande omschrijving van het begrip "schema" valt niet geheel samen met wat Schank en Abelson (1977) een script en een plan noemen. Een script is een bepaald type schema dat betrekking heeft op (de kennis van) een min of meer conventioneel patroon dat aan een sequentie van handelingen en gebeurtenissen ten grondslag ligt. Het is een "predetermined, stereotyped sequence of actions that defines a well-known situation" (Schank & Abelson, 1977, pag. 41). Een plan is een bepaald type schema dat vrij abstract is en bestaat uit "general information about how actors achieve goals" (Schank & Abelson, 1977, pag. 70).

Anderson (1978) omschrijft een schema als een abstracte kennisstructuur die uit een aantal componenten bestaat, die op een bepaalde wijze met el-

kaar zijn gerelateerd. Elke component bevat volgens hem een "slot" (open plaats) of variabele waaraan een bepaalde waarde kan worden toegekend. "A schema is an abstract description of a thing or event. It characterizes the typical relations among its components and contains a slot or placeholder for each component that can be instantiated with particular cases" (Pichert & Anderson, 1977, pag. 314). Uit de bovenstaande omschrijving blijkt dat Anderson c.s. praktisch dezelfde inhoud geven aan het begrip schema als Rumelhart c.s.

Volgens Rumelhart en Ortony (1977), Rumelhart en Norman (1978), Spiro (1977) en Anderson (1978) spelen schemata een essentiële rol bij het begrijpen en onthouden van (tekstuele) informatie.

Rumelhart en Ortony (1977) vatten begrijpen op als het selecteren van schemata en variabelen die passen bij de informatie die moet worden begrepen en het verifiëren of deze schemata inderdaad passen. Een schema past bij de inkomende informatie als die informatie kan worden geïnterpreteerd als een instantie van het concept dat het schema representeert. Een belangrijk gedeelte van het begrijpen heeft volgens Rumelhart en Ortony (1977) betrekking op het vinden van die schemata die het beste passen bij de inkomende informatie. Als eenmaal een geschikte verzameling van schemata is gevonden voor de inkomende informatie, dan worden de verschillende aspecten van de input (de constanten) gekoppeld aan de variabelen van de schemata en deze "geïnstantieerde" schemata vormen dan de interpretatie van de input. Verschillende schemata zullen volgens Rumelhart en Norman (1978) daarbij verschillende interpretaties van dezelfde input opleveren en verschillende kenmerken van de input zullen meer of minder belangrijk zijn als een functie van die interpretatie.

Volgens Rumelhart en Ortony (1977) constitueren schemata als het ware een soort theorie over die aspecten van de input die begrepen moeten worden. Het begrijpen lijkt volgens hen op een onderzoeksproces, waarbij evidentie wordt gezocht voor de bevestiging of falsificatie van een theorie. Schemata maken het mogelijk om afleidingen (inferenties) te maken en voorspellingen te doen omtrent aspecten van de input die nog niet of niet zijn geobserveerd. Als iemand bijvoorbeeld een tekst leest over een bezoek aan een dierentuin kan hij allerlei informatie afleiden en voorspellen. Als de geactiveerde schemata niet passen bij de input of de context waarbinnen de input wordt gepre-

senteerd, dan moeten nieuwe schemata worden geselecteerd. Wanneer geen geschikte schemata worden gevonden, dan blijft het begrijpen van de input beperkt tot fragmenten van die input (vgl. Bransford & Johnson, 1973).

Rumelhart en Norman (1978) veronderstellen dat schemata actieve kennis-eenheden zijn die de data, die via de zintuigen in het werkgeheugen zijn terecht gekomen, onderzoeken op relevantie voor hen. Wanneer een schema relevante data opmerkt, probeert het die data in haar structuur te integreren en voegt dan nieuwe data aan de reeds gegeven data toe. Andere schemata kunnen dan op deze nieuwe data reageren. Schemata worden dus door de data gestuurd in die zin dat ze reageren op voor hen belangrijke data. Van de andere kant activeren schemata zichzelf en geven ze richting aan het proces van begrijpen door gebruik te maken van hun interne conceptualisaties om nieuwe data toe te voegen. Volgens Rumelhart en Norman (1978) "each schema is data-driven and provides conceptually guided guidance to others" (pag. 44). In hun publicatie van 1977 spreken Rumelhart en Ortony van "bottom-up and top-down processing" (pag. 128).

De schematheorie impliceert volgens Anderson (1977) dat er bij het begrijpen van een tekst sprake is van een interactie tussen de kenmerken van de tekst (en de context) en de kennis van de lezer. "There is the "top-down" imposition of schemata...as well as the "bottom-up" thrust of data..." (pag. 417). Een bepaalde situatie interpreteren betekent volgens Anderson, Spiro en Anderson (1977) het vergelijken van de elementen van de situatie met de generieke kenmerken van de schematische kennisstructuur. Een boodschap begrijpen betekent volgens Anderson (1978) het plaatsen van een zodanige constructie op de boodschap dat een coherente formulering ontstaat van haar inhoud. Een coherente formulering betekent volgens hem een één-op-één relatie tussen de "slots" in het schema en de gegevenheden in de boodschap. Pichert en Anderson (1977) omschrijven het interpreteren van een boodschap als "matching the information in the message to the slots in a schema. The information entered into the slots is said to be subsumed by the schema" (pag. 314).

In de schematheorie wordt benadrukt dat de kennis waarover iemand beschikt een belangrijke factor is bij het begrijpen en onthouden van informatie. Omtrent de juistheid van dit algemene idee bestaat voldoende evidentie. Schank en Abelson (1977) geven een groot aantal voorbeelden waaruit

blijkt dat kennis van de wereld van essentieel belang is bij het begrijpen van zinnen en korte teksten. Een bekend voorbeeld is de zin: de agent stak zijn hand omhoog en stopte de wagen. Een ander voorbeeld in dit verband is het onderzoek van Bransford en Johnson (1973). Zij gaven proefpersonen een tekst zonder een duidelijk kader. Eén groep kreeg van tevoren een tekening waarin een mogelijke voorstelling van de tekstinhoud gegeven werd. Een andere groep kreeg alleen de tekst. Het resultaat was dat de eerste groep veel meer van de tekst begreep en onthield. Bovendien beoordeelde deze groep de tekst als gemakkelijk te interpreteren, terwijl de tweede groep de tekst juist moeilijk noemde. Het laatste betekent dat, ook al is de betekenis van alle woorden en zinnen bekend, dit nog niet voldoende is om de tekstuele informatie te begrijpen. Deze informatie moet ook binnen een bepaald kader geplaatst kunnen worden en dit kader is vaak gebaseerd op kennis van de werkelijkheid. Wanneer proefpersonen geen beroep kunnen doen op hun kennis van de wereld dan wordt het begrijpen van (tekstuele) informatie blijkaar in sterke mate bemoeilijkt.

Volgens de schematheorie varieert de belangrijkheid van een tekstelement wanneer lezers er toe worden gestimuleerd om schemata te activeren waarin een tekstelement een grote of minder grote rol speelt. Anderson c.s. hebben deze hypothese in enkele experimenten getest.

Anderson, Spiro en Anderson (1977) schreven twee teksten; een tekst over eten in een chic restaurant en een tekst over boodschappen doen in een supermarkt. Dezelfde achttien items over etenswaren, gekoppeld aan dezelfde personen, werden in dezelfde volgorde in de twee teksten genoemd. Voorspeld werd dat de proefpersonen die de restaurant-tekst kregen aangeboden de items over de etenswaren beter zouden onthouden dan de proefpersonen die de tekst over de supermarkt kregen aangeboden. Hierbij werd verondersteld dat een restaurantschema een meer precieze structuur voor etenswaren bevat dan een supermarktschema. Een tweede voorspelling was dat de proefpersonen die de restaurant-tekst kregen aangeboden de items over de etenswaren vaker aan de juiste personen zouden toewijzen. Een derde voorspelling was dat de proefpersonen die de restaurant-tekst lazen, beter in staat waren de volgorde van de items over de etenswaren te onthouden dan de andere proefpersonen. De eerste twee voorspellingen werden bevestigd; de derde voorspelling werd in één experiment bevestigd en in een ander experiment niet. Op grond van deze

resultaten concluderen Anderson, Spiro en Anderson (1977) dat de belangrjkheid van tekstelementen wordt bepaald door de schemata die tijdens het lezen worden geactiveerd.

In een onderzoek van Pichert en Anderson (1977) kregen proefpersonen de opdracht om twee teksten vanuit verschillende perspectieven - en zo werd verondersteld met verschillende schemata - te lezen. Eén tekst ging over een huis van een welgestelde familie. Het was een mooi, oud huis met enkele gebreken, zoals een lekkend dak, een vochtig souterrain en dergelijke. In het huis bevonden zich waardevolle dingen zoals een zeldzame collectie munten, een kleurentelevisie en dergelijke. Drie verschillende groepen proefpersonen kregen de opdracht om de verschillende tekstelementen naar belangrjkheid te scoren vanuit drie verschillende gezichtspunten: het gezichtspunt van een inbreker, het gezichtspunt van een huizenkoper en geen aangegeven gezichtspunt. De voorspelling, dat het gezichtspunt van de lezer bepaalde welke tekstelementen belangrjk werden gevonden en welke niet, werd bevestigd. In een daarop volgend experiment kregen andere proefpersonen één van de twee teksten aangeboden met de opdracht deze vanuit één van de drie gezichtspunten te lezen. Direkt na het lezen en na een week kregen de proefpersonen een free-recall tekst. Uit dit experiment bleek dat de tekstelementen die binnen een bepaald perspectief belangrjk waren, beter werden onthouden dan de minder belangrjke.

In een onderzoek van Anderson en Pichert (1978) kregen proefpersonen opnieuw de tekst over het huis van de welgestelde familie aangeboden met als opdracht deze tekst vanuit het perspectief van een huizenkoper en van een inbreker te lezen. Hierna kregen alle proefpersonen een free-recall test. Vervolgens kreeg de ene helft van de proefpersonen opnieuw een free-recall test vanuit hetzelfde perspectief. De andere helft kreeg als opdracht om van gezichtspunt te veranderen en zoveel mogelijk informatie op te schrijven vanuit het nieuwe gezichtspunt. Uit twee experimenten bleek dat de proefpersonen die van gezichtspunt veranderden significant meer informatie reproduceerden, die belangrjk was in het licht van het tweede perspectief, maar onbelangrjk in het licht van het eerste perspectief. Bovendien reproduceerden zij minder informatie die onbelangrjk was bij het tweede perspectief, maar belangrjk bij het eerste perspectief. Volgens Anderson en Pichert (1978) tonen deze resultaten aan dat herinneringsprocessen actieve, structu-

rerende processen zijn waarbij schemata een belangrijke rol spelen.

De bovengenoemde onderzoeken geven een beeld van het onderzoek dat in het kader van de schematheorie wordt uitgevoerd. We zullen nu enkele kritische opmerkingen plaatsen en een antwoord geven op de vraag in hoeverre de schematheorie als basis kan dienen voor de selectie van inferentiële leestaken.

De belangrijkste kritische opmerking die ten aanzien van de schematheorie of beter gezegd de schematheorieën gemaakt kan worden, heeft betrekking op haar geringe uitwerking tot nu toe. Een gedetailleerde kwantitatieve en kwalitatieve beschrijving van de rol die schemata spelen bij het begrijpen van tekstuele informatie ontbreekt (vgl. Glaser, Pellegrino & Lesgold, 1978). Deze opmerking impliceert in de eerste plaats dat het begrip "schema" scherper zal moeten worden omschreven (vgl. Hintzman, 1978; Hamaker, 1980). In de tweede plaats betekent deze opmerking dat preciezer zal moeten worden aangegeven hoe schemata werken. Het proces van de selectie van schemata en de vervanging van variabelen door constanten zal nauwkeurig moeten worden beschreven.

Een tweede opmerking betreft het feit dat met uitzondering van Rumelhart en Norman (1978) en Bobrow en Collins (1975) de schematheoretici tot nu toe te weinig aandacht besteden aan strategische kennis. Onder strategische of procedurele kennis verstaan we in dit verband de kennis van de lezer met betrekking tot de wijze waarop een leestaak moet of kan worden uitgevoerd. Verschillende schematheoretici besteden veel aandacht aan declaratieve kennis (knowing that) maar weinig aandacht aan de procedurele kennis (knowing how), die bij het leren begrijpend lezen van belang is (vgl. Gagné, 1977a; Glaser, Pellegrino & Lesgold, 1978; Hamaker, 1980; Kok, Boonman & Beukhof, 1980).

Een derde en laatste opmerking heeft betrekking op het feit dat in de schematheorie bij voorkeur aandacht wordt besteed aan het lezen van verhalen- de teksten en niet of slechts in beperkte mate aan informatieve teksten. De laatst genoemde teksten spelen in het onderwijs juist een belangrijke rol.

Hoewel de schematheoretische benadering interessante perspectieven biedt voor het onderzoek van het begrijpend leesproces is zij nog te weinig uitgewerkt om in dit onderzoek als basis te dienen voor de selectie van de inferentiële leestaken. De inferenties hebben in deze benadering bovendien in sterke mate betrekking op de kennis waarover de lezer al of niet beschikt en in mindere mate op de informatie die in de tekst gegeven is. Toch kunnen van-

uit deze benadering een aantal inferentiële leestaken worden geformuleerd zoals het afleiden van het thema van een tekst, het voorspellen van de afloop van een verhaal, het afleiden van de bedoeling van de actors in een verhaal, het afleiden van oorzaak-gevolg relaties. Uit het vervolg zal blijken dat één van de geselecteerde taken betrekking heeft op het afleiden van de bedoeling van de hoofdpersoon in een verhaal.

2.3. Inferenties en inferentieprocessen

In het model van Kintsch en Van Dijk (1978) spelen inferenties een minder belangrijke rol dan in de schematheorie van onder anderen Rumelhart en Ortony (1977). Volgens Kintsch en Van Dijk komt het inferentieproces vrij laat op gang en worden op het niveau van de microstructuur alleen proposities geïnferreed ten behoeve van de coherentie. Volgens Rumelhart en Ortony spelen inferenties bij het lezen van teksten een zeer belangrijke rol. Op basis van schemata zijn lezers volgens hen in staat om in een bepaalde tekst niet vermelde informatie af te leiden en op die manier verbindingen tot stand te brengen tussen schijnbaar niet verbonden proposities.

In het volgende zullen we verder ingaan op de problematiek van inferenties en inferentieprocessen. Eerst besteden we aandacht aan het begrip inferentie zoals dat door verschillende onderzoekers wordt omschreven. Daarna bespreken we enkele onderzoeken met betrekking tot het inferentieproces. Tot slot zullen we antwoord geven op de vraag in hoeverre de studie en het onderzoek op dit terrein een basis biedt voor de selectie van de leestaken.

De meeste onderzoekers hanteren het begrip inferentie noch in enge noch in ruime zin. Een enge interpretatie beperkt zich tot de logische inferenties, terwijl een ruime interpretatie betrekking heeft op alle niet expliciete informatie die naar aanleiding van een tekst wordt onthouden. Hoewel logische inferenties belangrijk zijn, zijn ze niet uitputtend voor de typen van inferenties die in de natuurlijke taal voorkomen. Anderzijds kan niet alle informatie die naar aanleiding van een tekst wordt onthouden als inferentiële informatie worden beschouwd. Een dergelijke opvatting is te ruim omdat hiermee foutief begrepen of onthouden informatie ten onrechte als inferentiële informatie kan worden opgevat én omdat daarmee het onderscheid tussen tekstuele informatie en reeds aanwezige kennis zou verdwijnen (vgl.

Sulin & Dooling, 1974).

Kintsch (1974) hanteert in eerste instantie een vrij beperkte omschrijving van inferenties. Hij breidt het begrip logische inferentie uit door dit te herinterpreteren binnen de "natural language logic". Dit betekent dat noties als kwantificatie en connectieven meer in overeenstemming worden gebracht met het gebruik in de natuurlijke taal. Kintsch (1974) formuleert ook een aantal relaties die in de logica niet of in andere vorm voorkomen zoals deel-geheel-, oorzaak-gevolg-, en categorie-relaties. Evenals bij presupposities en implicaties kan informatie die door deze relaties gegeven wordt, worden afgeleid. In een meer recente publicatie hanteert Kintsch (1977a) een ruimere omschrijving van inferenties. Naast inferenties die nodig zijn voor de noodzakelijke coherentie van een tekstbasis en inferenties van motieven voor acties, onderscheidt hij ook inferenties voor het samenvatten van een tekst. Macroproposities zoals generalisaties zijn volgens hem inferenties.

Bobrow (1975) definieert inferenties als impliciete feiten die worden afgeleid "from the initial set of explicit formulas according to some fixed rules of inference without interaction with the outside world" (pag. 16). Met deze omschrijving beperkt Bobrow zich voornamelijk tot logische inferenties. Schank (1975) spreekt over inferenties als het resultaat van processen die informatie afleiden "...that is likely to be true about a given input, but is not necessarily true" (pag. 167). Deze omschrijving staat een ruime interpretatie van het begrip inferentie toe. Brewer (1974) onderscheidt logische van pragmatische inferenties. De laatst genoemde inferenties worden ontleend aan verwachtingen gebaseerd op iemands kennis van de wereld en kunnen volgens Brewer worden geïdentificeerd met behulp van de "but not" test. Deze test houdt in dat een zin een andere zin impliceert als de impliciete zin in ontkennende vorm kan worden verbonden met de eerste zin door het woord "maar". De zin "The hungry python caught the mouse" impliceert "The hungry python ate the mouse" omdat "The hungry python caught the mouse, but did not eat it" een acceptabele zin is.

Frederiksen (1972) hanteert evenals Kintsch in eerste instantie een vrij beperkte omschrijving van inferenties. Hij onderscheidt inferentiële van elaboratieve processen. Bij inferentiële processen wordt volgens hem alleen noodzakelijk ware informatie afgeleid, zoals het afleiden van logische implicaties en presupposities. Bij elaboratieve processen wordt informatie af-

geleid die aan de (zwakke) eis voldoet dat deze niet in strijd is met de expliciete informatie. Het gaat hierbij om mogelijke, waarschijnlijke en plausibele informatie. In een meer recente publicatie, waarin een klassificatie van tekstuele inferenties wordt gepresenteerd, hanteert Frederiksen (1979) een zeer ruime omschrijving van inferenties. De klassificatie bestaat uit acht klassen van inferenties, die verder in 26 typen worden verdeeld. Enkele typen van inferenties zijn: lexical expansion, attribute inference, time inference, act inference, case inference, goal inference, theme inference, event inference, superordinate inference, causal inference, conditional inference, logical inference. De moeilijkheid van dit systeem bestaat volgens Goetz (1977) hierin dat alles wat naar aanleiding van een tekst wordt onthouden met een of andere propositie in de tekstbasis in verband wordt gebracht en dan wordt geklassificeerd als een inferentie. Het laatste betekent dat het begrip vrij inhoudsloos wordt.

In het vervolg zullen we het begrip inferentie in een vrij ruime betekenis hanteren. Dit wil zeggen dat we onder inferentie niet alleen die afgeleide informatie verstaan die noodzakelijk waar is, maar ook informatie, die mogelijk, waarschijnlijk of plausibel is. Informatie die het gevolg is van macroprocessen (Kintsch, 1977a) en het gebruik van schemata valt dus onder het begrip inferentie zoals hier opgevat.

Onderzoeken van onder anderen Bransford et al. (1972), Brewer (1974) en Potts (1972) hebben overtuigend aangetoond, dat inferenties optreden bij het lezen (of horen) en onthouden van zinnen.

Bransford, Barclay en Franks (1972) gaven hun proefpersonen de volgende zin (1) "Three turtles rested on a floating log and a fish swam beneath them". Vervolgens kregen de proefpersonen de zin (2) "Three turtles rested on a floating log and a fish swam beneath it" met daarna de opdracht te bepalen of zin (2) overeenkwam met zin (1). De meeste proefpersonen waren van mening dat ze zin (2) ook de eerste keer hadden gehoord. Volgens Bransford c.s. is het door elkaar halen van zin (1) met zin (2) te wijten aan een inferentieproces. De verwarring van beide zinnen is het gevolg van kennis die mensen hebben van spatiële relaties en die hen in staat stelt zin (2) als een implicatie van zin (1) te zien. Bransford c.s. konden aantonen dat deze resultaten niet verklaard konden worden door een geringe aandacht van de proefpersonen voor de exacte formulering van de zinnen.

Johnson, Bransford en Solomon (1973) toonden aan dat proefpersonen ook inferenties maken over objecten die niet in zinnen voorkomen. Ze gaven proefpersonen een zin als (1) "John was trying to fix the birdhouse. He was pounding the nail when his father came out to watch and to help him do the work". Een controlegroep kreeg de volgende zin (2) "John was trying to fix the birdhouse. He was looking for the nail when his father came out to watch him and to help him do the work". Later werd aan beide groepen gevraagd of de volgende zin overeenkwam met wat ze oorspronkelijk hadden gehoord (3) "John was using the hammer to fix the birdhouse when his father came out to watch him and to help him do the work". De resultaten wezen uit dat de eerste groep veel meer dan de controlegroep meende dat zin (3) overeenkwam met wat ze oorspronkelijk hadden gehoord.

Brewer (1974) toonde aan dat proefpersonen geneigd waren zich inferenties te herinneren die waarschijnlijke, hoewel niet logisch noodzakelijke, consequenties waren van de zinnen die werden aangeboden. Zinnen als (1) "The hungry python caught the mouse" werden vaak gereproduceerd als (2) "The hungry python ate the mouse". Brewer verklaart dit als gevolg van het feit dat zin (2) een acceptabele implicatie vormt van zin (1).

Potts (1972) en Barclay (1973) toonden aan dat wanneer proefpersonen een reeks van zinnen kregen aangeboden zoals (1) "Bill is stronger than Tom" en (2) "Tom is stronger than John" zij ten onrechte zinnen herkenden als "Bill is stronger than John". Het is duidelijk dat zin (3) logisch uit de zinnen (1) en (2) kan worden afgeleid. In het natuurlijke taalgebruik komen logische inferenties vaak in een andere vorm voor. Clark (1977) geeft het volgende voorbeeld: "Max lives in New York. Moritz is crazy too" (pag. 254).

Kintsch (1974) en Singer (1976; 1977) hebben verschillende onderzoeken uitgevoerd met het doel vast te stellen wanneer inferenties plaatsvinden. Inferenties kunnen op verschillende momenten voorkomen nl. tijdens het proces van begrijpen of tijdens het herinneringsproces. Verder is de mogelijkheid niet uitgesloten dat inferenties zowel tijdens het proces van begrijpen als tijdens het herinneringsproces voorkomen.

Kintsch (1974) voerde drie experimenten uit om aan te tonen dat inferenties tijdens het leesproces plaatsvinden. Bij de eerste twee experimenten werden zowel korte als langere teksten gebruikt. Van elke tekst werd een expliciete en een impliciete versie geconstrueerd. Het belangrijkste verschil

tussen deze twee versies bestond hierin dat bij de impliciete versie één propositie (bij de korte teksten) of meerdere proposities was (waren) wegge-
laten. Een voorbeeld van de versies van een korte tekst is het volgende:

Explicit: A carelessly discarded burning cigarette started a fire. The fire
destroyed many acres of virgin forest.

Implicit: A burning cigarette was carelessly discarded. The fire destroyed
many acres of virgin forest. (Kintsch, 1974, pag. 155).

Eén groep proefpersonen kreeg de expliciete versie (22 experimentele tek-
sten); een andere groep kreeg de impliciete versie (22 experimentele tek-
sten). Na elke tekst kregen de proefpersonen een testzin aangeboden met als
opdracht zo snel mogelijk te beslissen of de aangeboden zin waar of onwaar
was op grond van de expliciete of impliciete informatie in de tekst. De
testzin bestond uit de propositie(s) die in de impliciete versie moest(en)
worden afgeleid. De testzin van het bovenstaande voorbeeld is: "The discar-
ded cigarette caused the fire". Uitgaande van de veronderstelling dat infe-
renties tijdens het proces van begrijpen plaatsvinden en dat de betreffende
proefpersonen dus tijdens het lezen eenzelfde propositionele representatie
construeren, voorspelde Kintsch dat er geen verschillen tussen de latentie-
tijden zouden bestaan voor de impliciete en de expliciete versie. Deze voor-
spelling werd niet bevestigd. De impliciete versie leidde tot meer fouten in
de verificatietaak en tot langere latentietijden dan de expliciete versie.
Kintsch verklaart dit als gevolg van de invloed van de oppervlakte-represen-
tatie die direct na het lezen nog aanwezig is en waarvan de proefpersonen
met de expliciete versie het meest profiteren. In een volgend experiment,
dat qua design was verbeterd, kreeg elke proefpersoon 15 teksten achter el-
kaar aangeboden. Na een tijdsinterval van 15 minuten werden de testzinnen
aangeboden. Bij dit experiment werden geen verschillen gevonden tussen de
impliciete en expliciete versie. Niet duidelijk is echter wat de invloed is
van het vergeten van de informatie bij de uitvoering van de verificatietaak.
Eenzelfde resultaat werd bij het derde experiment gevonden bij tijdsinter-
vallen van 20 minuten en 48 uur. De resultaten van deze experimenten moeten
met de nodige voorzichtigheid worden gehanteerd vanwege methodologische on-
volkomenheden en de soms grote verschillen tussen de expliciete en implicie-
te versie van de teksten.

Singer (1976) vond dat proefpersonen kort na het lezen van zinnen als

"The small girl spent the gleaming penny" sneller in staat waren om een parafrase van dergelijke zinnen te verifiëren dan een inferentie ("The penny was new"). Singer (1977) voerde drie experimenten uit waarbij de proefpersonen een herkenningstaak of een verificatietask moesten uitvoeren. De proefpersonen kregen korte teksten van drie zinnen aangeboden. In de expliciete versie was de eerste zin hetzelfde als de testzin ("The sailor swept the floor with the broom"). In de impliciete versie bestond de eerste zin uit een implicatie van de testzin ("The sailor swept the floor in the cabin"). In het eerste experiment bleek dat de verificatietijd voor de impliciete items (inferenties) langer was dan voor de expliciete items. In het tweede experiment werden de proefpersonen 6 seconden na het horen van de tekst getest. Opnieuw bleek dat de expliciete items sneller werden geverifieerd dan de impliciete items. In het derde experiment werd de testzin van de expliciete versie vervangen door een parafrase. Ook nu bleek dat de expliciete items i.c. de parafrasen sneller werden geverifieerd dan de inferenties. Op grond van deze resultaten concludeert Singer dat minstens een deel van het inferentieproces tijdens het herinneren moet plaatsvinden. Hoewel alternatieve verklaringen voor de gevonden resultaten mogelijk zijn, lijkt deze conclusie in het algemeen niet onjuist. Goetz (1977) merkt op dat de vraag of inferenties tijdens het proces van begrijpen of tijdens het herinneringsproces voorkomen niet juist gesteld is. Aangenomen moet worden dat het inferentieproces zowel tijdens het begrijpen als tijdens het herinneren kan optreden.

Verschillende onderzoekers zoals Haviland en Clark (1974), Thorndike (1976) en Goetz (1977) hebben experimenten uitgevoerd met het doel vast te stellen onder welke condities inferenties worden gegenereerd (vgl. Manoliu-Dabija & Schipper, 1980; Van Oostendorp & Den Uyl, 1980).

Volgens Haviland en Clark (1974) wordt een zin sneller begrepen wanneer deze past binnen een bepaalde context. Als er geen duidelijke context beschikbaar is, moet de lezer "backward inferences" maken naar een voorgaande context, hetgeen inspanning en tijd kost. In een drietal experimenten toonden Haviland en Clark aan dat wanneer een bepaalde zin ("target sentence") wordt voorafgegaan door een zin waarin expliciet de "gegeven" informatie van de betreffende zin ("target sentence") wordt vermeld deze sneller wordt begrepen dan wanneer er geen duidelijk antecedent voor de gegeven informatie

beschikbaar is. Zo werd de "target sentence" (1) "The beer was warm" sneller begrepen wanneer deze werd voorafgegaan door de zin (2) "We got some beer out of the thrunk" dan door de zin (3) "Andrew was especially fond of beer". De verbinding tussen de zinnen (3) en (1) kost meer tijd, omdat een extra inferentiële stap nodig is.

Thorndike (1976) schreef vier teksten waarin telkens twee "eigenaardige" zinnen voorkwamen. De eerste zin, de "target sentence", was van dien aard dat deze aanleiding gaf tot het genereren van enkele inferenties. Een voorbeeld van een dergelijke zin met mogelijke inferenties is:

"The hamburger chain owner was afraid his love for french fries would ruin his marriage.

The hamburger chain owner got his french fries for free.

The hamburger chain owner's wife didn't like french fries.

The hamburger chain owner was very fat." (Thorndike, 1976, pag. 439). Verderop in de tekst volgde een tweede zin, de "continuation sentence", die in de experimentele conditie de functie had om de plausibiliteit van één mogelijke inferentie te vergroten. Een voorbeeld van een dergelijke zin is: "The hamburger chain owner decided to join weight-watchers in order to save his marriage". In de controle conditie volgde een "continuation sentence", die de functie had om de plausibiliteit van de verschillende inferenties niet te beïnvloeden. Een voorbeeld van een dergelijke zin is: "The hamburger chain owner decided to see a marriage counselor in order to save his marriage". In het eerste experiment toonde Thorndike aan dat proefpersonen in de experimentele conditie vaker de juiste inferentie genereerden dan de neutrale of foutieve inferentie. In het tweede experiment kregen de proefpersonen een herkenningstest met de instructie om bevestigend te reageren op die items die expliciet in de tekst werden vermeld en ontkennend op die items die konden worden afgeleid. Uit de gegevens bleek dat de proefpersonen in de experimentele conditie de meeste fouten maakten bij die inferenties die juist waren. Op grond van deze resultaten concludeert Thorndike dat mensen zowel de expliciet geformuleerde informatie van een tekst onthouden evenals de inferenties die nodig zijn voor de integratie van die informatie (vgl. Bower, 1976).

Uitgaande van het reeds jaren bekende gegeven (vgl. Bartlett, 1932) dat belangrijke elementen van een tekst beter worden onthouden dan onbelangrijke,

voorspelde Goetz (1977) dat lezers eerder een inferentie maken bij belangrijke dan bij niet belangrijke tekstelementen. Bovendien voorspelde hij dat zeer saillante premissen in een tekst van invloed zijn op het vóórkomen van een inferentie. Goetz stelde zes verhalende teksten samen van ongeveer 500 woorden en varieerde de relevantie van een inferentie evenals de saillantie van de premisse. In de helft van de tekstversies werd de inferentie expliciet geformuleerd. De beide voorspellingen werden in twee experimenten bevestigd. Evenals in het onderzoek van Thorndike (1976) bleek verder dat proefpersonen ten onrechte van mening waren dat een belangrijke inferentie expliciet in de tekst was geformuleerd.

Het bovenstaande overzicht met betrekking tot inferenties en inferentieprocessen is beperkt en geeft slechts een indruk van het onderzoek en de theorievorming op dit gebied. Uit het onderzoek zoals hier vermeld blijkt, dat inferenties vóórkomen en een belangrijke rol spelen bij het lezen en onthouden van informatie. Uit de laatst vermelde experimenten blijkt verder dat onderzoekers zich in toenemende mate bezig houden met de vraag hoe inferentieprocessen werken en welke variabelen hierop van invloed zijn. Om deze vraag op een adequate wijze te beantwoorden, is een algemene theorie over inferentieprocessen noodzakelijk. Volgens Kintsch en Van Dijk (1978) ontbreekt een dergelijke theorie. Dit laatste betekent voor het onderhavige onderzoek dat de keuze van inferentiële leestaken (nog) niet gebaseerd kan worden op een theorie over inferentieprocessen.

Uit deze en de vorige paragraaf blijkt dat de keuze van inferentiële leestaken nog niet gebaseerd kan worden op een uitgewerkte en enigszins getoetste theorie over begrijpend lezen en inferentieprocessen. Dit impliceert dat de taken die geselecteerd worden en de instrumenten die hiervoor ontwikkeld worden nog niet voldoende kunnen worden gevalideerd, althans wat betreft de begripsvaliditeit. Aan de eis, die Farr (1969) terecht stelt aan de constructie van meetinstrumenten op het gebied van begrijpend lezen nl. dat de keuze van de te meten taken gebaseerd moet zijn op een theorie van begrijpend lezen, kan op dit moment slechts in beperkte mate worden voldaan.

2.4. Selectie en specificatie van de leestaken

2.4.1. Selectie van de taken

Naast de eis om de keuze van de leestaken zoveel mogelijk te baseren op een theorie van begrijpend lezen, werd in de beginfase van het onderhavige onderzoek nog een aantal criteria geformuleerd waaraan de te selecteren taken moesten voldoen. Deze criteria waren:

- a. de taken moeten primair betrekking hebben op het begrijpen van teksten en niet primair op het begrijpen van woorden of zinnen;
- b. de taken moeten betrekking hebben op het afleiden van informatie uit een tekst;
- c. de taken moeten niet te veel een beroep doen op feitelijke, specifieke kennis;
- d. de taken moeten - voor zover dat valt te voorzien - relevant zijn voor leesonderwijs en ook via leesonderwijs zijn te beïnvloeden;
- e. de taken moeten - voor zover dat valt te bepalen - geen overlap vertonen;
- f. de taken moeten met behulp van klassikale, schriftelijke tests te meten zijn.

De eerste drie criteria werden opgesteld naar aanleiding van een eerder uitgevoerde literatuurstudie met betrekking tot begrijpend lezen (vgl. Aarnoutse et al., 1977). Het eerste criterium (a) is vooral gebaseerd op de overweging dat begrijpend lezen in de schoolsituatie primair een kwestie is van het begrijpen van teksten en niet van losse woorden of zinnen. Het tweede criterium (b) is gebaseerd op de veronderstelling dat het afleiden van informatie een essentieel aspect vormt van begrijpend lezen. Het derde criterium (c) is slechts globaal geformuleerd en duidt aan dat de rol van feitelijke kennis beperkt moet blijven. Feitelijke kennis zal evenals gestructureerde kennis (schema-kennis) altijd een rol spelen bij het begrijpen van teksten.

De laatste drie criteria werden opgesteld naar aanleiding van het doel van het onderhavige onderzoek. Criterium (d) geeft aan dat bij een onderzoek naar het effect van drie experimentele programma's - het derde doel van het onderzoek - taken moeten worden geselecteerd waarvan redelijkerwijs kan worden aangenomen dat ze van belang zijn voor leesonderwijs en ook via leesonderwijs zijn te beïnvloeden. Criterium (e) houdt rechtstreeks verband met

het vierde doel van het onderzoek nl. het vaststellen van de realiteitswaarde van de onderscheiding van taken. Dit criterium is slechts globaal geformuleerd en geeft aan dat taken waarvan vooraf kan worden voorspeld dat ze nauw met elkaar samenhangen niet voor selectie in aanmerking komen. Taken als bijvoorbeeld het afleiden van de hoofdgedachte en het afleiden van het thema van een informatieve tekst komen tezamen niet voor selectie in aanmerking omdat ze te veel in elkaars verlengde liggen (vgl. Otto, 1977). Het laatste criterium (f) houdt rechtstreeks verband met het tweede doel van het onderzoek nl. het vaststellen van de betrouwbaarheid en validiteit van klassikaal afneembare tests voor begrijpend lezen. Dit criterium impliceert dat de te selecteren taken zo exact mogelijk moeten kunnen worden omschreven; niet alleen wat betreft het uit te voeren gedrag maar ook wat betreft het object i.c. de tekst.

Zoals uit de vorige paragrafen is gebleken, zijn theorieën over begrijpend lezen nog niet voldoende uitgewerkt en getoetst om als basis te dienen voor de selectie van de leestaken. Wel was het mogelijk om een aantal inferentiële leestaken uit het model van Kintsch en Van Dijk (1978) en de schematheorie van Rumelhart en Ortony (1977) af te leiden en te formuleren. Ook konden enkele taken worden geformuleerd vanuit de studie en het onderzoek van inferentieprocessen. Mogelijke taken waren onder andere het afleiden van logisch geldige informatie en het afleiden van informatie, die mogelijk, waarschijnlijk of plausibel is.

Bij de selectie van de leestaken werd nagegaan welke taken voldeden aan de bovenvermelde criteria. Hierbij werd duidelijk dat het laatste criterium (f) nl. de ontwikkeling van klassikale, schriftelijke tests bij verschillende taken problemen opriep. Uiteindelijk werden de volgende taken geselecteerd: het afleiden van de hoofdgedachte uit een informatieve tekst (de taak Hoofdgedachte), het afleiden van de bedoeling van de hoofdpersoon uit een fictieve, verhalende tekst (de taak Bedoeling), en het trekken van conclusies uit logische redeneringen (de taak Redeneren).

De taak Hoofdgedachte heeft betrekking op het selecteren en ordenen van tekstuele informatie op basis van een centraal thema (inductieve inferentie). Deze taak werd vooral gekozen omdat ze past binnen het model van Kintsch en Van Dijk (1978). Bovendien wordt in de meeste Amerikaanse en in

enkele Nederlandse leesprogramma's expliciet aandacht besteed aan deze taak. Ook in het Wisconsin Design vormt deze taak (main idea) een belangrijke component van begrijpend lezen (vgl. Aarnoutse et al., 1976).

De taak Bedoeling heeft betrekking op het interpreteren van informatie over handelingen binnen een consistent kader, zó dat motieven voor handelingen kunnen worden afgeleid (intentionele inferentie). Deze taak werd in eerste instantie gekozen naar aanleiding van een publicatie van Kintsch (1977a) waarin deze uitdrukkelijk spreekt over inferenties van motieven voor handelingen. Hoewel deze taak niet expliciet wordt genoemd in de voorlopige lijst van doelstellingen van het leesonderwijs (vgl. Aarnoutse et al., 1976), bleek uit een analyse van enkele leesprogramma's en tests dat vaak vragen worden gesteld die betrekking hebben op het vinden van de bedoeling van de hoofdpersoon.

De taak Redeneren past binnen een engere definitie van het begrip "inferentie" omdat het hierbij gaat over het afleiden van logisch geldige informatie (logische inferentie). Deze taak werd vooral gekozen naar aanleiding van de publicaties van Frederiksen (1972; 1979) en Kintsch (1974). In veel Amerikaanse leesprogramma's wordt aan deze taak expliciet aandacht besteed. In de voorlopige lijst van doelstellingen van het leesonderwijs (vgl. Aarnoutse et al., 1976) is deze taak geplaatst binnen de subcategorie "kritisch lezen". In het Wisconsin Design vormt deze taak (reasoning) één van de componenten van begrijpend lezen.

De drie bovengenoemde taken voldoen - voor zover dat valt te voorzien - aan de eerder gestelde criteria. Wat betreft de taak Redeneren kan men zich afvragen of deze taak van belang is voor het onderwijs in begrijpend lezen. Uitgaande van het idee dat kritisch lezen een belangrijk aspect vormt van begrijpend lezen, zijn we geneigd deze vraag bevestigend te beantwoorden. Naar onze mening dient het begrijpend leesonderwijs in de lagere school de nodige aandacht te besteden aan bepaalde vormen van logisch redeneren.

2.4.2. Specificatie van de taken

In verband met de specificatie van de taken werd een literatuurstudie uitgevoerd. Deze studie had tot doel de taken nader te omschrijven en voor zover mogelijk theoretisch te funderen. In het volgende worden voor elke taak de belangrijkste bevindingen van de literatuurstudie vermeld. Daarna

volgt een specificatie van de betreffende taak. Voor meer informatie verwijzen we naar Deelrapport 4 (Aarnoutse et al., 1979).

Het afleiden van de hoofdgedachte

De problematiek van de hoofdgedachte van een tekst werd vanuit twee verschillende invalshoeken bestudeerd. In de eerste plaats werd nagegaan op welke wijze het begrip hoofdgedachte in de didactische literatuur wordt omschreven en welk onderzoek en ontwikkelingswerk op dit terrein is verricht. In de tweede plaats werd de problematiek benaderd vanuit de psycholinguïstische literatuur. Nagegaan werd welke kenniscomponenten van belang zijn bij de constructie van een globale tekstrepresentatie en welke factoren en regels verantwoordelijk zijn voor een dergelijke representatie.

Wat betreft de omschrijving van het begrip hoofdgedachte werden publicaties van onder anderen Gates (1947), Bond en Wagner (1960), Chapman (1971), Jolly (1974), Harris en Sipay (1977), Malmquist en Brus (1974), Drop en De Vries (1977) en Leistra (1979) bestudeerd. Geconstateerd werd dat in het merendeel van deze publicaties het begrip hoofdgedachte en daarmee samenhangend het begrip thema slechts in algemene termen wordt omschreven en dat een nauwkeurige analyse en beschrijving van deze begrippen meestal ontbreekt. Wat betreft onderzoek en ontwikkelingswerk op het gebied van de hoofdgedachte bleek dat Otto en Barrett (1968) en Van Blaricom en White (1975; 1976) verschillende onderzoeken en ook ontwikkelingswerk hebben verricht dat duidelijke aanwijzingen bevat voor de opzet en constructie van een test en een programma. Vermeldenswaard is in dit verband de omschrijving van Otto en Barrett (1968) van het begrip hoofdgedachte. Volgens hen moet een adequate formulering van de hoofdgedachte twee elementen bevatten nl. (a) een verwijzing naar een algemeen onderwerp zoals bijvoorbeeld in de formulering "Vogels bouwen nesten" en (b) een beperking van het algemene onderwerp door de verwijzing naar de specifieke inhoud van de tekst zoals bijvoorbeeld in de zin "Vogels bouwen nesten op verschillende plaatsen".

Vanuit de psycholinguïstische literatuur was het mogelijk om langs indirecte weg een bijdrage te leveren aan de verheldering van de begrippen hoofdgedachte en thema. Uitgaande van de theorie van Kintsch en Van Dijk (1978) is het bepalen of afleiden van de hoofdgedachte van een tekst een essentieel aspect van de vorming van een coherente macrostructuur. Bij het be-

palen van de hoofdgedachte van een informatieve tekst is sprake van de constructie van een nieuw concept (superconcept) op basis van andere concepten. Het thema van een tekst kan worden omschreven als het begrip waarop uiteindelijk alle informatie is terug te voeren. Het thema drukt uit waar de tekst over gaat, terwijl de hoofdgedachte aangeeft wat de tekst over het thema zegt. Een duidelijk aanwijsbare grens tussen de beide begrippen ontbreekt.

Alvorens een operationele definitie te geven van het afleiden van de hoofdgedachte, noemen we eerst de belangrijkste criteria waaraan de teksten moeten voldoen. Deze criteria geven aan het begrip hoofdgedachte, zoals hier gehanteerd, een bepaalde inhoud en voorkomen onduidelijkheid in deze (vgl. Anderson & Pichert, 1978). De belangrijkste criteria of kenmerken zijn:

- a. de teksten moeten informatief zijn. Dit wil zeggen dat ze beschrijvend moeten zijn en niet fictief, overredend of directief;
- b. de teksten moeten over één bepaald onderwerp handelen;
- c. de teksten moeten een aantal uitspraken over het onderwerp bevatten, die tezamen één bepaalde gedachte tot uitdrukking brengen. Deze uitspraken dienen de functie te hebben van illustrerende bijzonderheden, voorbeelden of uitwerkingen van één bepaalde niet expliciet geformuleerde gedachte.

Een essentieel onderdeel van het afleiden van de hoofdgedachte is het selecteren van een thema. Het thema geeft antwoord op de vraag waar de tekst over gaat. (Van Dijk (1978) noemt dit het globale thema). In de tekst worden over dit thema een aantal uitspraken gedaan die, in vergelijking met een meer algemene uitspraak, specifiek van aard zijn en één gemeenschappelijk kenmerk hebben. De hoofdgedachte is nu die uitspraak die in meer algemene termen weergeeft wat de specifieke uitspraken tezamen tot uitdrukking brengen.

Een operationele definitie van het afleiden van de hoofdgedachte van een informatieve tekst kunnen we nu als volgt formuleren: vind een uitspraak over het thema van de tekst zó dat de verschillende mededelingen fungeren als illustrerende bijzonderheden, voorbeelden of uitwerkingen hiervan. Essentieel voor het afleiden van de hoofdgedachte is dus dat een gemeenschappelijk kenmerk van bepaalde uitspraken in de tekst wordt opgespoord. Dit impliceert een proces waarbij de leerling over specifieke uitspraken moet generaliseren (vgl. de generalisatie-regel van Van Dijk (1978)). Met de term "afleiden" wordt in dit verband niet bedoeld dat de leerlingen de hoofdgedachte zelf

genereren of formuleren. Het is de bedoeling dat de leerlingen de hoofdgedachte van een tekst uit een viertal alternatieven selecteren. Uit onderzoek van Van Blaricom en White (1975) blijkt dat het selecteren van de hoofdgedachte uit een aantal alternatieven gemakkelijker is dan het genereren van de hoofdgedachte. Bij het laatste wordt naast het begrijpend lezen van de tekst ook een beroep gedaan op de bekwaamheid van de leerling de hoofdgedachte adequaat te formuleren. De term "afleiden" wordt hier in een meer beperkte betekenis gebruikt.

Het afleiden van de bedoeling

In verband met de specificatie van deze taak werden publicaties van onder anderen Miller en Johnson-Laird (1976), Abelson (1975), Schank en Abelson (1977), Rumelhart (1977b), Kintsch (1977a) en Thorndike (1977) bestudeerd.

Het onderkennen of afleiden van een handelingsmotief is een bekwaamheid die de nodige kennis veronderstelt. Miller en Johnson-Laird (1976) gaan uit van het concept INTEND voor de beschrijving van de kennis die ten grondslag ligt aan het interpreteren van intentionele handelingen. Aan dit concept onderscheiden zij drie componenten nl. een handelende persoon, een doel en een plan om dit doel te bereiken. Hoewel dit concept slechts beperkt van omvang is, vormt het wel de basis voor de interpretatie van intentionele handelingen. Volgens Abelson (1975) en Schank en Abelson (1977) begrijpen mensen handelingssequenties in termen van doelgerichte handelingen. Voor de interpretatie van handelingen (acts) wordt een onderliggend plan geconstrueerd. Hoewel een plan een variabele manier is om een doel te bereiken, onderhouden de onderdelen van een plan een zekere structurele relatie. Kennis van deze structurele patronen is daarom van belang bij het onderkennen of afleiden van een handelingsmotief.

Bij het afleiden van de bedoeling van de hoofdpersoon uit een verhalen- de tekst is nog een andere vorm van kennis van belang nl. de kennis die mensen hebben van de structuur van verhalende teksten. Rumelhart (1977b) geeft de volgende grondstructuur van een verhalende tekst: er is een gebeurtenis waardoor het doel dat de hoofdpersoon (Protagonist) wil bereiken gemotiveerd wordt (Cause). De hoofdpersoon probeert (Try) dit doel te bereiken, al of niet via een serie subdoelen. Het uiteindelijke resultaat (Outcome) is bepalend voor het al of niet bereikt zijn van het gestelde doel. Volgens Kintsch

(1977a) bestaan eenvoudige verhalen uit een reeks episodes. Elke episode bestaat uit een expositie, een complicatie en een oplossing. De complicatie van een episode kan weer uit een hele episode bestaan. Tenslotte heeft een verhaal vaak een moraal aan het einde. Thorndike (1977) onderscheidt de volgende hoofdelementen aan een verhaal: de "setting" (de begintoestand), het "theme" (een of meer gebeurtenissen die de hoofdpersoon overkómen en een doel dat hiervan het gevolg is), de "plot" (het gedeelte van het verhaal waarin de handelingen worden verricht die tot het gestelde doel leiden) en de "resolution" (de ontknoping: het resultaat van het handelen van de hoofdpersoon).

Rumelhart (1977b), Kintsch (1977a) en Thorndike (1977) hebben een aantal onderzoeken uitgevoerd waaruit blijkt dat kennis van de structuur van eenvoudige verhalen een rol speelt bij het begrijpen hiervan. Mandler en Johnson (1977) toonden aan dat kinderen van de eerste en vierde klas van de lagere school deze kennis reeds gebruiken.

Alvorens een operationele definitie te geven van het afleiden van de bedoeling noemen we eerst de criteria waaraan de teksten moeten voldoen. Deze criteria zijn:

- a. de teksten moeten een verhalend karakter hebben en niet beschrijvend of beschouwend zijn;
- b. de teksten bevatten tenminste één handelend persoon;
- c. in de tekst wordt een aantal handelingen beschreven die zich binnen een bepaalde plaats- en tijdsruimte afspelen;
- d. de tekst bevat een aantal aanwijzingen voor de bedoeling van de hoofdpersoon. Deze aanwijzingen kunnen expliciet zijn (bijvoorbeeld door het gebruik van werkwoorden als "willen" en "van plan zijn"), maar ook impliciet;
- e. de tekst bevat een resultaat van het handelen.

De vraag die de leerlingen moeten beantwoorden is die naar het handelingsmotief van de hoofdpersoon. Dit motief (de bedoeling) staat niet expliciet in de tekst. De vraag naar de bedoeling is een vraag naar de uiteindelijke bedoeling, voor zover die uit de tekst is af te leiden. Wanneer de tekst dus aanwijzingen bevat voor een dieperliggende bedoeling, dan is deze het juiste antwoord op de vraag.

De gestelde taak kunnen we operationeel als volgt omschrijven: leid een

handelingsmotief af uit een verhalende tekst waarin een handelingssequentie wordt beschreven als onderdeel van de uitvoering van een plan zonder dat expliciet wordt aangegeven wat het doel hiervan is. Met de term "afleiden" wordt in dit verband niet bedoeld dat de leerlingen de bedoeling van de hoofdpersoon zelf genereren of formuleren. De leerlingen moeten het meest adequate motief uit een viertal alternatieven selecteren. Evenals bij de hoofdgedachtetaak wordt de term "afleiden" hier in een meer beperkte betekenis gebruikt.

Het afleiden van logisch geldige informatie (redeneren)

Voor de specificatie van deze taak werden publicaties van onder anderen Toulmin (1958), Ennis (1976), Piaget (1963), Kohnstamm (1934; 1952), Hill (1961), Roberge (1970), Ennis en Paulus (1965) en Carroll (1975) bestudeerd.

Volgens Toulmin (1958) voldoen slechts weinig argumentaties aan de formele eisen van de logische redenering. De formele logica is volgens hem geen geschikt middel bij het beoordelen van argumentaties. Eist de formele logica dat redeneringen geldig zijn en de conclusies ondubbelzinnig waar of onwaar, de argumentatietheorie kent slechts deugdelijke redeneringen en aanvaardbare conclusies. In tegenstelling tot de argumentatietheorie richt de formele logica zich uitsluitend op de geldigheid van de redeneervorm en abstraheert daarmee van de inhoud van de premissen. Deze abstractie van de inhoud heeft geleid tot enkele heroriëntaties in de logica. Ennis (1976) geeft een overzicht van verschillende logische systemen en van geldige en ongeldige redeneervormen.

Volgens Piaget (1963) zijn kinderen pas vanaf het elfde levensjaar in staat om formeel te redeneren. Vanaf dit jaar gaat zich het logisch systeem als systeem ontwikkelen. Uit onderzoek van Kohnstamm (1934), Hill (1961) en Ennis (1976) blijkt dat kinderen jonger dan elf jaar al in staat zijn tot logisch redeneren. Uit onderzoek van Ennis (1976) blijkt dat er geen sprake is van de ontwikkeling van een logisch systeem, maar van een geleidelijke ontwikkeling van logische redeneervormen. De ontwikkeling van de verschillende redeneervormen verloopt niet parallel; sommige redeneervormen worden eerder beheerst dan andere. Roberge (1970) vindt geen verschillen in ontwikkeling tussen enkele redeneervormen uit verschillende logische systemen.

Uit onderzoek van Ennis en Paulus (1965), Roberge (1970) en Carroll

(1975) blijkt dat geldige redeneervormen zoals modus ponens (Alle A zijn B, n is een A, dus n is een B) en modus tollens (Alle A zijn B, n is geen B, dus n is geen A) gemakkelijker zijn voor leerlingen van de lagere school dan ongeldige redeneervormen zoals conversie (Alle A zijn B, n is een B, dus n is een A) en inversie (Alle A zijn B, n is geen A, dus n is geen B). Uit deze onderzoeken blijkt verder dat slechts een beperkt aantal redeneervormen voor het onderhavige onderzoek, dat betrekking heeft op leerlingen van het vierde leerjaar, in aanmerking kunnen komen. Op grond van deze onderzoeken en de redeneertoets van het Wisconsin Design for Reading Skill Development is gekozen voor de geldige redeneervormen modus ponens en modus tollens en de ongeldige redeneervormen conversie en inversie. De syllogistische vormen uit de predicaatlogica en de conditionele vormen uit de propositielogica zijn bij de verdere uitwerking door elkaar gebruikt.

Alvorens een operationele definitie te geven van de taak Redeneren, noemen we eerst een aantal kenmerken waaruit deze bestaat. De taak bestaat uit de opdracht een conclusie - vraag te beantwoorden over een tekst, die een major en een minor bevat. Een major is een algemene uitspraak die categorisch of hypothetisch kan zijn. Een categorische major drukt uit dat een eigenschap wordt toegekend aan een gehele klasse. Deze uitspraak kent dus een eigenschap B toe aan alle leden van klasse A. Een categorische major heeft de vorm: Alle A zijn B. Een hypothetische major heeft de vorm: Als A, dan B. Elk type major heeft twee termen nl. een A-term en een B-term.

Naast een major bevat een tekst ook een minor. Een minor is een specifieke uitspraak, die een eigenschap toekent aan één lid van de klasse, waarover de major een algemene uitspraak doet. Er zijn vier typen minors: de minor kan de A-term van de major bevestigen (+) of ontkennen (-), of de B-term van de major bevestigen of ontkennen.

Naast een major en een minor bestaat een tekst uit een aantal andere zinnen. Deze zijn voor de redenering niet van belang. Door deze zinnen krijgen de relevante zinnen (de major- en de minorzin) een context. De tekst moet zodanig worden geconstrueerd, dat niet is na te gaan of hetgeen in de tekst beschreven is empirisch waar is of niet.

Bij elke tekst wordt een conclusie - vraag gesteld. Deze vraag heeft betrekking op de conclusie die uit de major en de minor getrokken kan worden. Bij deze vraag moeten de leerlingen beoordelen of het antwoord op deze vraag

positief of negatief is, dan wel of de tekst te weinig informatie bevat om deze vraag met zekerheid te kunnen beantwoorden.

De gestelde taak kunnen we operationeel als volgt omschrijven: gegeven een tekst waarin een major en een minor voorkomen, leid hieruit een conclusie af. Met de term "afleiden" wordt in dit verband niet bedoeld dat de leerlingen de conclusie zelf moeten genereren of formuleren. De leerlingen moeten het juiste antwoord uit drie alternatieven selecteren. De term "afleiden" wordt ook bij deze taak in een meer beperkte betekenis gebruikt.

Kintsch en Van Dijk (1978) veronderstellen dat inferenties plaatsvinden op basis van algemene en contextuele kennis. Volgens Rumelhart en Ortony (1977) worden inferenties gegenereerd op basis van georganiseerde kennisstructuren of schemata. In het onderhavige onderzoek wordt verondersteld dat niet alleen gestructureerde kennis maar ook strategische of procedurele kennis een rol kan spelen bij het afleiden van informatie. Onder strategische kennis verstaan we in dit verband de kennis van de lezer met betrekking tot de wijze waarop een bepaalde leestaak kan worden uitgevoerd. Opgemerkt moet worden dat Rumelhart en Norman (1978) uitgaan van de veronderstelling dat gestructureerde kennis (schemakennis) ook procedureel van aard is.

Als we nu de drie bovengenoemde taken vergelijken wat betreft de gestructureerde kennis waarop bij de uitvoering van deze taken een beroep kan worden gedaan, dan vallen een aantal verschillen op. De taak Hoofdgedachte en de taak Bedoeling verschillen vooral in twee opzichten van elkaar. In de eerste plaats is de rol van een eventuele conventionele tekststructuur bij de taak Hoofdgedachte van weinig belang, terwijl deze bij de taak Bedoeling van invloed kan zijn. In de tweede plaats verschillen deze taken in de aard van de gestructureerde kennis waarop een beroep kan worden gedaan. Bij de taak Bedoeling is kennis omtrent het conventionele verloop van doelgerichte handelingen waarschijnlijk van groot belang. Als deze kennis in onvoldoende mate beschikbaar is, dan zal deze taak waarschijnlijk niet goed kunnen worden uitgevoerd. Bij de taak Hoofdgedachte wordt eveneens een beroep gedaan op gestructureerde kennis; deze kennis heeft voornamelijk betrekking op concepten. Vermoed wordt dat bij de taak Hoofdgedachte de gestructureerde kennis niet zo'n grote tol zal spelen als bij de taak Bedoeling. Tussen de

taak Redeneren en de beide andere taken bestaan eveneens enkele verschillen. Bij de taak Redeneren zal gestructureerde kennis waarschijnlijk een beperkte rol spelen. Deze kennis zal zelfs interfereren met de taak. Zo verwachtte Piaget (1963) dat kinderen moeite zouden hebben met een logische redenering wanneer de inhoud van de premissen strijdig was met hun kennis van de werkelijkheid. Dit effect is door verschillende onderzoekers gevonden (vgl. Carroll, 1975). In de tweede plaats verschilt de taak Redeneren van de beide andere taken in die zin dat een meer stringente analyse van de componenten van de taak kan worden gegeven. De nadruk op de redeneervorm maakt het mogelijk om de taak nauwkeurig te specificeren in termen van goed gedefinieerde begrippen.

Uit het voorgaande kan worden afgeleid dat de drie geselecteerde taken verschillen wat betreft de gestructureerde kennis waarop bij de uitvoering van deze taken een beroep wordt gedaan. Uitgaande van de veronderstelling dat er een sterk verband bestaat tussen het gedrags- en inhoudsaspect van een taak, verwachten we dat de drie taken ook verschillen wat betreft de strategische kennis waarop bij de uitvoering van de taken een beroep kan worden gedaan. Dit betekent dat we geen transfer-effect verwachten van de programma's die gericht zijn op een betere uitvoering van de verschillende taken.

Zoals uit hoofdstuk I blijkt, staan in dit onderzoek drie vragen centraal nl. de vraag naar het effect van drie taakgerichte programma's op het gebied van begrijpend lezen, de realiteitswaarde van de onderscheiding van drie taken met betrekking tot begrijpend lezen en de vraag naar de betrouwbaarheid en validiteit van vijf tests voor begrijpend lezen. Dit hoofdstuk bevat het resultaat van de literatuurstudie die in verband met deze vragen werd verricht.

In par. 3.1. besteden we aandacht aan programma's (curricula) en benaderingen van begrijpend lezen en aan onderzoek naar het effect van strategieën en experimentele programma's. Deze paragraaf wordt afgesloten met een verantwoording van de opzet van de programma's die in dit onderzoek worden ontwikkeld en getoetst.

In verband met de vraag naar de realiteitswaarde van de onderscheiding van de drie inferentiële leestaken wordt in par. 3.2. eerst een beschrijving gegeven van de vaardigheden die een aantal auteurs aan begrijpend lezen onderscheiden. Daarna volgt een overzicht van het onderzoek naar vaardigheden op het gebied van begrijpend lezen.

Par. 3.3. begint met een korte beschrijving van de verschillende methoden voor het meten van begrijpend lezen. Daarna wordt aandacht besteed aan de twee typen tests die in dit onderzoek centraal staan en aan onderzoek dat in verband met de validiteit van deze tests van belang is. Op basis van de bestudeerde onderzoeksliteratuur worden tenslotte enkele verwachtingen geformuleerd omtrent mogelijke verbanden tussen de te ontwikkelen tests en maten voor technisch lezen, begrijpend lezen, woordenschat en (non-verbale) intelligentie.

3.1. Programma's voor begrijpend lezen en hun effect

Zoals reeds eerder vermeld, heeft dit onderzoek onder andere tot doel het effect vast te stellen van drie experimentele programma's (leereenheden). Het antwoord op de vraag naar het effect van deze programma's is vooral didactisch van belang omdat hierdoor een indicatie kan worden verkregen van de waarde i.c. het effect van taakgericht onderwijs op het gebied van begrij-

pend lezen. In par. 3.1.1. besteden we in het kort aandacht aan programma's en benaderingen van begrijpend lezen in de Verenigde Staten en meer uitvoerig aan twee controverse benaderingen en hun effect. Volledigheidshalve volgt in par. 3.1.2. een overzicht van programma's of curricula die in Nederland zijn of worden ontwikkeld. In par. 3.1.3. bespreken we een aantal onderzoeken naar het effect van strategieën en experimentele programma's op het gebied van begrijpend lezen. Par. 3.1.4. geeft in het kort een verantwoording van de opzet van de programma's die in dit onderzoek worden ontwikkeld en getoetst.

3.1.1. Programma's voor begrijpend lezen in de Verenigde Staten

Zoals reeds in hoofdstuk I is vermeld, werd in een eerder onderzoek (vgl. Aarnoutse et al., 1976) een aantal leesprogramma's bestudeerd en nagegaan op welke wijze en op grond van welke theorieën, concepten en wetenschappelijke gegevens de doelstellingen van deze programma's waren vastgesteld, geordend en geformuleerd. In totaal werden zes Amerikaanse en twee Nederlandse programma's geselecteerd en geanalyseerd nl. Individually Prescribed Instruction-Reading Program, ontwikkeld door het Learning Research and Development Center te Pittsburgh; Scott Foresman Reading Systems, ontwikkeld door Scott Foresman and Company te Glenview; Wisconsin Design for Reading Skill Development, ontwikkeld door het Wisconsin Research and Development Center for Cognitive Learning te Madison; Instructional Objectives Exchange - Reading, ontwikkeld door het Instructional Objectives Exchange te Los Angeles; SOBAR/Mastery: an evaluation tool, ontwikkeld door het Center for the Study of Evaluation te Los Angeles; System for Pupil and Program Evaluation and Development, ontwikkeld door het New York State Education Department te Albany; Correct leren lezen, ontwikkeld door een werkgroep onder leiding van Van der Schoot; en Met inzicht leren lezen, ontwikkeld door een werkgroep onder leiding van Van Eeuwijk-Rijnen.

De bovengenoemde programma's waren niet alle van dezelfde aard. De eerste drie en de laatste twee programma's zijn geheel of gedeeltelijk uitgewerkte curricula met handleidingen, werkboeken en dergelijke. De overige programma's zijn doelstellingen- en toetssystemen die bij bestaande curricula kunnen worden gebruikt. Het Scott Foresman programma is een typisch voorbeeld van wat in de Verenigde Staten een "basal reader" wordt genoemd, een pro-

gramma dat decennia geleden is ontwikkeld, regelmatig wordt herzien, en dat materialen en aanwijzingen bevat voor het leesonderwijs in leerjaar 1 t/m 6. (Volgens Spache (1972) worden in 80% van de scholen "basal readers" gebruikt in de leerjaren 3 t/m 6).

Uit de analyse van de bovengenoemde programma's bleek dat de indeling, bepaling en sequentiëring van de doelen van de verschillende programma's slechts in beperkte mate was gebaseerd op theoretische inzichten en gegevens van empirisch onderzoek en voor een belangrijk deel berustte op voorwetenschappelijke kennis en ervaring. Bovendien bleek dat bij de ontwikkeling van de verschillende programma's slechts in beperkte mate gebruik was gemaakt van evaluatie-onderzoek. Met uitzondering van het Wisconsin Design kon over het effect van de programma's of onderdelen daarvan niets of nauwelijks iets worden gerapporteerd.

Hoewel de geanalyseerde programma's slechts een fractie vormen van het totale bestand aan leesprogramma's in de Verenigde Staten en Nederland geven de resultaten van de analyse waarschijnlijk toch een duidelijke indicatie van de stand van zaken wat betreft de theorievorming, het onderzoek en het ontwikkelingswerk op het gebied van het (voortgezet) leesonderwijs in 1975-1976. Hoe het begrijpend leesonderwijs de facto in Amerikaanse scholen wordt gerealiseerd, vermeldt Durkin (1978-1979). Zij vond dat slechts een gering percentage van de geobserveerde instructietijd daadwerkelijk besteed wordt aan begrijpend leesonderwijs.

Het eerder uitgevoerde onderzoek is in zoverre eenzijdig dat alleen programma's zijn geanalyseerd, waarin de verschillende doelen (algemene, intermediaire en concrete) expliciet worden vermeld. Hoewel dit selectie criterium om praktische redenen noodzakelijk was, was het gevolg dat met uitzondering van het Scott Foresman programma vooral programma's zijn geanalyseerd met de "skill-approach" als uitgangspunt. Dit is echter niet de enige benadering van leesonderwijs.

In het volgende zullen we aan de hand van publicaties van Spache en Spache (1973) en Kamm (1978) in het kort beschrijven welke benaderingen van leesonderwijs in de Verenigde Staten gangbaar zijn. Daarna besteden we aandacht aan twee controversiële benaderingen op het gebied van (begrijpend) lezen en aan onderzoek naar het effect daarvan.

Spache en Spache (1973) onderscheiden vier benaderingen of methoden om

kinderen te leren lezen nl. de "basal reader approach", de "individualized approach", de "linguistic approaches" en de "language experience approach". De "basal reader approach" kan men globaal omschrijven als die benadering waarbij aan de hand van een uitgewerkt programma - bijvoorbeeld Scott Foresman Reading Systems - klassikaal leesonderwijs wordt gegeven. De "individualized approach" houdt zo veel mogelijk rekening met de individuele verschillen tussen leerlingen en probeert aan te sluiten bij de ontwikkeling en behoeften van de leerlingen. Het Wisconsin Design is een voorbeeld van deze benadering. Binnen de "linguistic approaches" onderscheiden Spache en Spache drie richtingen nl. de fonologische, structuralistische en psycholinguïstische richting. De fonologische richting legt sterk de nadruk op de relatie tussen letter- en klankpatronen bij het aanvankelijk lezen. De structuralistische richting beklemtoont het belang van syntactische aspecten tijdens het lezen. Volgens Lefevre (1964) moeten kinderen vooral via voorlezen geattendeerd worden op het bestaan en de functie van allerlei zinspatronen. De psycholinguïstische richting komt volgens de auteurs vooral tot uiting in de publicaties van Goodman (1968). De "language experience approach" is volgens Spache en Spache eigenlijk een methode voor het aanvankelijk leesonderwijs. Deze benadering gaat uit van de spontane uitingen van het kind en beklemtoont de nauwe relatie tussen taal, taalontwikkeling en lezen (vgl. Stauffer, 1970). De benadering vertoont veel overeenkomst met wat in Nederland de globaal methode wordt genoemd (vgl. Dumont, 1971; Aarnoutse et al., 1975).

Kamm (1978) onderscheidt vier benaderingen om lezen te onderwijzen nl. de "skill-based approach", de "holistic approach", de "psycholinguistic approach" en de "common sense approach". De "skill-based approach" gaat volgens haar uit van belangrijke deelvaardigheden op het gebied van lezen en beschouwt deze als componenten van het hoger geordende leesproces. De "holistic approach" beklemtoont volgens Kamm dat lezen een integraal proces is dat niet in componenten uiteengelegd en onderwezen kan worden. De "psycholinguistic approach" benadrukt de toepassing in het leesonderwijs van wat bekend is omtrent de rol van grafisch-fonologische, syntactische en semantische informatie tijdens het leesproces. De "common sense approach" is volgens Kamm een alternatieve benadering die de sterke kanten van de verschillende benaderingen tracht te integreren.

Uit het bovenstaande blijkt dat de benaderingen die Spache en Spache on-

derscheiden slechts gedeeltelijk overeenkomen met de methoden van Kamm. Een mogelijke verklaring hiervoor is dat de benaderingen van Spache en Spache vooral betrekking hebben op het aanvankelijk lezen (leerjaar 1 en 2) en de methoden van Kamm op het voortgezet lezen (leerjaar 3 tot en met 6). Opvallend in beide publicaties is verder dat de benaderingen slechts in algemene termen worden omschreven. Voor een overzicht van de resultaten of beter gezegd het gebrek aan resultaten van het groots opgezette onderzoek van Bond en Dijkstra (1967) naar de effectiviteit van verschillende methoden in het aanvankelijk leesonderwijs verwijzen we naar Deelrapport 1 (Aarnoutse et al., 1975).

Sedert een aantal jaren bestaat in de Verenigde Staten een controverse tussen voorstanders van de "skill-approach" en van de "holistic approach" op het gebied van (begrijpend) lezen. Voorstanders van de "skill-approach" zijn onder anderen Davis (1972), Samuels (1976), Otto (1977) en Kamm (1978). Voorstanders van de "holistic approach" zijn onder anderen Goodman (1970; 1973; 1976), Smith (1978) en Vacca (1980).

Zoals we in hoofdstuk I hebben gezien, is het Wisconsin Design gebaseerd op de "skill-approach". In dit programma wordt uitgegaan van een groot aantal vaardigheden verdeeld over gebieden of aspecten van lezen en geordend in componenten (strands). Elke component bevat een aantal vaardigheden en doelen die nauw aan elkaar zijn gerelateerd en die in niveaus (A tot en met C) zijn gerangschikt. De procedure van het Wisconsin Design bestaat in grote lijnen hierin dat eerst wordt vastgesteld welke vaardigheden en doelen de leerlingen voor de verschillende gebieden en componenten reeds hebben verworven of bereikt. Voor de vaststelling van de leesvaardigheidsontwikkeling van de leerlingen wordt onder meer gebruik gemaakt van doelstellinggerichte tests. Vervolgens worden groepen van leerlingen samengesteld die dezelfde vaardigheid of combinatie van vaardigheden trachten te verwerven. Na een pre-test krijgen deze groepen gedurende een bepaalde periode onderwijs waarbij het accent ligt op het leren verwerven van een strategie voor de uitvoering van de betreffende vaardigheid. Bovendien krijgen de leerlingen de nodige oefeningen (opgaven), die in subgroepen of individueel worden gemaakt. Tenslotte volgt een post-test (een parallelvorm van de pre-test) om vast te stellen of de betreffende vaardigheid/doelstelling is verworven respectievelijk bereikt. Toepassing van de geleerde vaardigheid vindt meest-

al in andere vakken of leergebieden plaats.

Deze korte beschrijving van de procedure van het Wisconsin Design komt voor een deel overeen met de definitie die Vacca (1980) geeft van "subskill instruction". Hij omschrijft deze vorm van leesonderwijs als volgt: "Instruction centered around a set of comprehension subskills arranged in a logical sequence and taught in a prescribed manner. The teacher's main objectives are to: 1) identify appropriate skill needs of students through informal diagnosis, 2) teach the relevant subskills, 3) provide practice in those subskills and 4) provide for application of those subskills" (pag. 515). Opgemerkt moet worden dat in het Wisconsin Design vaak afzonderlijk aandacht wordt besteed aan de ontwikkeling van één vaardigheid, terwijl in andere programma's die eveneens op de "skill-approach" zijn gebaseerd meerdere vaardigheden gedurende eenzelfde leereenheid aan de orde komen. Deze programma's gaan vaak uit van een centrale tekst waarna een aantal vaardigheden worden geoefend (vgl. de leergang "Met inzicht leren lezen" van Van Eeuwijk-Rijnen et al. (1973) en de leergang "Begrijpend lezen" van het project Onderwijs en Sociaal Milieu (Nuy, 1980)).

In zijn verantwoording van de "skill-approach" hanteert Otto (1977) voornamelijk praktische argumenten. De meest rechtstreekse manier om begrijpend lezen te onderwijzen, bestaat volgens Otto in het onderwijzen van de belangrijke (deel)vaardigheden. Vaardigheden hebben we volgens hem nodig om het begrijpend leesonderwijs op een gerichte en systematische wijze te realiseren. Zonder vaardigheden is het niet goed mogelijk om te diagnostiseren, te individualiseren of te differentiëren. Vaardigheden mogen echter niet geïsoleerd worden onderwezen en moeten steeds beschouwd worden als een middel voor de optimalisering van het begrijpend leesproces. "I see skills and the teaching of skills as the substance and the means for sharpening the process of developing reading comprehension. Skills must be identified, taught, and applied in context. In perspective skills can be the vehicle for moving children to independence in their reading comprehension" (Otto, 1977, pag. 196). Volgens Otto behoeft er geen tegenstelling te bestaan tussen de "skill-approach" en de "holistic approach". Naar zijn mening moet er niet alleen aandacht besteed worden aan de belangrijke vaardigheden van begrijpend lezen maar ook aan het leesproces zelf dat de simultane toepassing vereist van een reeks van vaardigheden. Beide benaderingen vullen elkaar aan.

Athey (1977) is evenals Otto van mening dat de "skill-approach" een belangrijke bijdrage kan leveren aan de optimalisering van het begrijpend leesonderwijs. Zij wijst echter op twee problemen die inherent zijn aan deze benadering. Het eerste probleem heeft betrekking op de validiteit van de vaardigheden die in programma's, die op deze benadering zijn gebaseerd, worden onderscheiden. Volgens Athey bestaat er over de vaardigheden die in het Wisconsin Design voor begrijpend lezen zijn geselecteerd tot dusverre slechts "some consensual validation" (pag. 235). Het tweede probleem heeft betrekking op de interactie van de vaardigheden en de rol daarvan tijdens het begrijpend leesproces. Volgens Athey krijgt dit probleem dat onder andere betrekking heeft op de transfer van de geleerde vaardigheden naar het begrijpend lezen in de dagelijkse situatie te weinig aandacht in de "skill-approach". Ook Samuels (1976) en Kamm (1978) wijzen nadrukkelijk op de noodzaak van integratie van de geleerde vaardigheden.

Kamm (1978) heeft een onderzoek gedaan naar het effect van de invoering van het Wisconsin Design in een schooldistrict in de Verenigde Staten. In totaal waren 80 lagere scholen met ongeveer 40.000 leerlingen bij dit onderzoek betrokken. In 1971-1972 werd decoderen (word attack) als eerste element van het Wisconsin Design in de betreffende scholen ingevoerd; in 1972-1973 volgde studievvaardigheden (study skills) en in 1974-1975 begrijpend lezen (comprehension). Vanaf 1971-1972 werden elk jaar bij de leerlingen van leerjaar 4 en 6 vier subtests afgenomen van een norm-referenced test nl. de Comprehensive Tests of Basic Skills. De subtests hadden betrekking op woordenschat, begrijpend lezen, het gebruik van "reference materials", en het gebruik van "graphic materials". Vanaf 1974 werden bij alle leerlingen van de betreffende scholen (K-6) de tests van het Wisconsin Design voor decoderen en studievvaardigheden afgenomen. Vanaf 1976 werden ook de tests van het Wisconsin Design op het gebied van begrijpend lezen afgenomen.

Uit een vergelijking van de scores behaald in 1971-1972 en 1976-1977 op de vier subtests van de Comprehensive Tests of Basic Skills bleek dat de prestaties van de leerlingen van leerjaar 4 en 6 aanzienlijk waren verbeterd. Uit een vergelijking van de scores behaald in 1974 en 1977 op de tests van het Wisconsin Design voor decoderen en studievvaardigheden bleek dat de prestaties van de leerlingen in sterke mate waren gestegen. Het percentage leerlingen van leerjaar 3 dat bezig was met de verwerving van de decodeer-

vaardigheden op niveau D steeg van 47.1% in 1974 tot 75.0% in 1977. Het percentage leerlingen van leerjaar 4 dat bezig was met de verwerving van studievaardigheden op niveau E steeg van 26.6% in 1974 tot 50% in 1977. Uit een vergelijking van de scores behaald in 1976 en 1977 op de tests van het Wisconsin Design voor begrijpend lezen bleek eveneens dat de prestaties van de leerlingen in sterke mate waren gestegen. Het percentage leerlingen van leerjaar 4 dat bezig was met de vaardigheden voor begrijpend lezen op niveau E steeg van 9.6% in 1976 tot 54.0% in 1977 (zie voor de betekenis van niveaus par. 1.2. en 3.2.1.).

De resultaten van dit onderzoek moeten met de nodige voorzichtigheid worden gehanteerd omdat niet duidelijk is welke variabelen (de inservice training, de schoolorganisatie, het Wisconsin Design, het testonderzoek en dergelijke) van invloed zijn geweest op de resultaten (vgl. Tierney & Cunningham, 1980). Uit de resultaten mag in ieder geval niet worden afgeleid dat de "skill-approach" een betere benadering is om kinderen te leren lezen. De resultaten geven echter wel een duidelijke indicatie dat invoering van het Wisconsin Design van invloed is op de leesprestaties van de leerlingen.

De voorstanders van de "holistic approach" beklemtonen dat lezen een taalproces is dat niet in componenten of vaardigheden verdeeld kan worden zonder het in wezen aan te tasten. Lezen is in hoge mate één activiteit of handeling en elke poging om het in delen uiteen te leggen of te onderwijzen betekent een ontkenning van de eigen aard van lezen. Volgens Goodman (1972), Smith (1978) en Vacca (1980) moet lezen als één integraal proces worden onderwezen waarbij het begrijpen van de betekenis van de boodschap van de schrijver centraal staat. Kinderen leren volgens deze auteurs vooral lezen door te lezen. De rol van de leerkracht dient beperkt te blijven tot begeleiden en stimuleren. "The teacher is not so much a source of wisdom in sound reading instruction as a guide and aid, monitoring the learner's progress, offering help when a hang-up is detected, stimulating interest in reading, helping him find relevant, worthwhile materials to read, and offering continuous encouragement" (Goodman, z.j., pag. 4).

Uit het bovenstaande zal duidelijk zijn dat voorstanders van de "holistic approach" grote vraagtekens plaatsen bij de "skill-approach". Zij trekken de validiteit van vaardigheden in twijfel evenals het sequentieel arrangeren van vaardigheden of deelvaardigheden. Ook de notie van de beheersing van

vaardigheden en de meting daarvan verwerpen ze. Volgens hen moet zowel bij het aanvankelijk lezen (leerjaar 1 en 2) als bij het voortgezet lezen (leerjaar 3 tot en met 6) het gehele leesproces en de functie van dit proces nl. het vatten van de betekenis centraal staan. Het onderwijzen van allerlei vaardigheden en deelvaardigheden werkt eerder belemmerend dan faciliterend.

Vacca (1980) omschrijft "holistic instruction" als volgt: "Instruction centered around strategies designed to engage students actively in reading through a constant interaction with print. The teacher's main objectives are 1) to involve students in reading as an entity in itself and 2) to emphasize the students' results of reading through discussion and/or writing activities" (pag. 514). In zijn onderzoek hanteert Vacca vier strategieën die volgens hem een operationalisatie vormen van de bovengenoemde definitie. Deze strategieën zijn:

- a. Uninterrupted Sustained Silent Reading. Deze strategie van Hunt (1970) stimuleert leerlingen de belangrijkste gedachten uit geschreven informatie te selecteren. De leerkracht probeert de leerlingen ervan bewust te maken dat lezen één geheel is door vragen te stellen als: Hoe heb je vandaag gelezen? Heb je veel gelezen?
- b. Request. Deze strategie van Manzo (1969) bestaat hierin dat de leerkracht en de leerling elkaar om beurten vragen stellen over delen van een tekst. Als leerlingen in staat zijn voorspellingen te doen over de afloop van een verhaal, lezen ze de tekst verder door en nemen ze deel aan de discussie.
- c. Radio Reading. Deze strategie van onder anderen Greene (1979) maakt gebruik van hardop lezen om het belang van de communicatie van ideeën te beklemtonen. Bij deze procedure leest een leerling voor, terwijl de andere leerlingen luisteren. Daarna volgt een discussie.
- d. Directed Reading - Thinking Activity. Deze strategie van Stauffer (1969) bestaat hierin dat de leerlingen eerst de inhoud van een tekst moeten voorspellen op basis van de titel. Vervolgens lezen de leerlingen de tekst waarbij na elk deel de leerkracht de volgende vragen stelt: Wat denk jij? Waarom denk je zo? Kun je het bewijzen? Als blijkt dat de leerlingen hun predicties bijstellen en hun reacties verantwoorden, mogen ze op een zeker moment de tekst verder aflezen. Daarna volgt een discussie.

Het gemeenschappelijke van de bovengenoemde strategieën bestaat hierin dat in tegenstelling tot de "skill-approach" het leesproces zo veel mogelijk als één proces wordt onderwezen. Wat betreft het effect van deze strategieën wijzen Tierney en Cunningham (1980) in hun overzicht op enkele onderzoeken waaruit blijkt dat de strategieën van Manzo en Stauffer in het algemeen effect hebben. Toch is er volgens hen te weinig onderzoek naar deze strategieën gedaan "to prompt any general or differential suggestions regarding either the construction or implementation of these strategies" (pag. 33). Ook Vacca (1980) is van mening dat er weinig bekend is over het effect van de "holistic approach".

Vacca (1980) deed een onderzoek naar het effect van de "holistic approach" en de "skill-approach" op het gebied van begrijpend lezen. Nagegaan werd wat het effect was van beide benaderingen op de (begrijpende) leesprestaties en leesattitudes van 34 leerlingen uit de leerjaren 7 tot en met 9. De geselecteerde leerlingen vertoonden een duidelijke achterstand op het gebied van begrijpend lezen ondanks een normale begaafdheid. De eerste groep van 17 leerlingen kreeg gedurende drie weken leesonderwijs volgens de bovengenoemde strategieën van de "holistic approach". Zeven tutores gaven deze vorm van leesonderwijs aan groepjes van 2 à 3 leerlingen. De tweede groep van 17 leerlingen kreeg gedurende dezelfde drie weken leesonderwijs volgens de "skill-approach". Ook nu gaven zeven tutores deze vorm van leesonderwijs aan groepjes van 2 à 3 leerlingen. Na de eerste drie weken werd een norm-referenced test voor begrijpend lezen afgenomen evenals een leesattitudeschaal. Na deze periode volgde een tweede periode van drie weken waarbij de twee groepen wisselden van benadering. De eerste groep kreeg leesonderwijs volgens de "skill-approach" en de tweede groep volgens de "holistic approach". Op het einde van de tweede periode werd een parallelvorm van de eerder gebruikte norm-referenced test afgenomen en opnieuw dezelfde leesattitudeschaal. Uit de gegevens van dit experiment bleek dat er geen significant verschil was tussen de beide benaderingen. Er werden ook geen volgorde- of interactie-effecten gevonden. Wat betreft de leesattitudes werden evenmin significante verschillen gevonden tussen beide benaderingen. Wel werd er een significant interactie-effect gevonden tussen de beide benaderingen en de volgorde van presentatie. Vacca veronderstelt dat de gekozen norm-referenced test waarschijnlijk niet gevoelig is om effecten op korte termijn

vast te stellen. Hoewel deze verklaring wellicht juist is, ligt een andere verklaring nog meer voor de hand nl. dat het doel waarop beide programma's waren gericht - verbetering van het begrijpend lezen in het algemeen - te omvattend is om in zo'n korte tijd te bereiken. Het stellen van een dergelijk omvattend doel is echter inherent aan de "holistic approach".

In zijn kritiek op de opvattingen van Goodman (1972) en Smith (1973) wijst Samuels (1976) op verschillende onderzoeken waaruit blijkt dat het leren van een complexe vaardigheid zoals lezen wordt vergemakkelijkt door eerst een aantal deelvaardigheden te leren. Volgens hem is de beheersing van een complexe vaardigheid of taak afhankelijk van de beheersing van een aantal deelvaardigheden van lagere orde. "While agreeing with the critics of the sub-skill approach, that too much emphasis can be placed on these subordinate skills, the critics probably are in error in failing to recognize the importance of subskills in the developmental sequence of skill attainment" (Samuels, 1976, pag. 176). Volgens Samuels behoeft er geen tegenstelling te bestaan tussen de "holistic approach" en de "skill-approach". Beide benaderingen erkennen volgens hem dat er (deel)vaardigheden bestaan. De "skill-approach" begint met kleinere eenheden en werkt vervolgens naar grotere en meer complexe eenheden, terwijl de "holistic approach" met een grotere eenheid begint en vervolgens aandacht besteedt aan kleinere eenheden.

Een goed voorbeeld van integratie van beide benaderingen op het gebied van begrijpend lezen geven Collins en Smith (1980). Volgens hen kan het proces van begrijpend lezen in drie fasen worden onderwezen. In de eerste fase leest de leerkracht een verhaal of een tekst voor, terwijl hij al lezende hardop hypothesen formuleert, informatie zoekt voor de bevestiging of verwerping van de hypothesen, de betekenis van woorden op grond van de context achterhaalt, de relatie tussen zinnen of paragrafen opspoort etc. De volgende fase bestaat hierin dat de leerkracht de leerlingen aanmoedigt om de vaardigheden die de leerkracht in de eerste fase demonstreerde zelf te gebruiken terwijl ze hardop lezen. In de derde fase lezen de leerlingen een tekst stil en krijgen ze tijdens het lezen van de tekst schriftelijke opdrachten en vragen met de bedoeling de geleerde vaardigheden toe te passen. Na het lezen van de tekst worden eveneens vragen gesteld en beantwoord.

3.1.2. Programma's voor begrijpend lezen in Nederland

Een belangrijke pionier op het gebied van het begrijpend leesonderwijs in Nederland is ongetwijfeld Van Veen geweest die in 1929 een verhandeling schreef over de plaats en betekenis van het stil lezen in de lagere school. In deze publicatie - mededeling 8 van het Nutsseminarium - beklemtoont hij het belang van het stil lezen als methode om het zelfstandig denken te oefenen en te stimuleren. De leerlingen moeten zich volgens hem leren concentreren op de tekst en gegevens die verspreid in de tekst liggen, leren combineren (oefening van het concentratie- en combinatievermogen). Een geschikte vorm waarin het stil lezen volgens Van Veen kan worden beoefend, is de invulvorm (cloze procedure).

In latere publicaties van het Nutsseminarium - mededelingen 23 en 24 van Van der Hoeve, Kohnstamm en Van Veen (1933; 1935) - wordt het stil lezen of het zelfstandig verwerken van leesstof vooral gezien als een belangrijk middel voor de toetsing van de intelligentie en voor de selectie van kinderen voor wat tegenwoordig het voorbereidend wetenschappelijk onderwijs heet. In dit verband hebben Kohnstamm en zijn medewerkers vooral aandacht besteed aan de opbouw en de inhoud van de teksten en aan de aard en de diagnostische waarde van de vragen. Vanuit hun denkpsychologische achtergrond kenden zij veel waarde toe aan de zogenaamde gebonden en vrije combinatievragen.

Het stil lezen als vorm van begrijpend lezen heeft in het lager onderwijs vooral ingang gevonden dankzij de stilleesmethoden "Stillezen en zelf doen", ontwikkeld door Kohnstamm en zijn medewerkers, "Op weg naar denkend lezen" en "Denkend lezen", ontwikkeld door Langeveld en Palland (vgl. Deen, 1969). De methode "Stillezen en zelf doen" verscheen in 1938 en was bestemd voor het 4e, 5e en 6e leerjaar. In de methode komen veel teksten voor die Kohnstamm en zijn medewerkers in eerdere onderzoeken hebben gebruikt. De vragen worden in vier soorten onderscheiden nl. lexicologische vragen, oriënteringsvragen, gebonden combinatievragen en vrije combinatievragen. Over het algemeen zijn de teksten lang en het aantal vragen zo groot dat betwijfeld moet worden of de leerkrachten die deze methode gebruikten voldoende aandacht hebben kunnen besteden aan het leren van geschikte oplossingsmethoden voor het vinden van de juiste antwoorden.

De bovengenoemde stilleesmethoden van Kohnstamm en zijn medewerkers of leerlingen zijn niet alleen van betekenis geweest voor het begrijpend lees-

onderwijs tijdens en na de tweede wereldoorlog maar hebben ook invloed uitgeoefend op de ontwikkeling van leesmethoden of -programma's die in de zestiger jaren verschenen. In tegenstelling tot de methoden van Kohnstamm c.s. besteden de leesmethoden uit de zestiger jaren meer aandacht aan allerlei leesvormen zoals groepslezen, forumlezen, kringlezen en dergelijke. In de volledig herziene uitgave van Kohnstamm's methode met als titel "Stillezen en er wat mee doen" (vgl. Bossers, 1974) wordt aan groepslezen veel waarde toegekend.

In ons land zijn de laatste jaren enkele analyses uitgevoerd van een aantal veel gebruikte methoden, programma's of leergangen voor begrijpend lezen. Daandels (1974) heeft een analyse uitgevoerd van acht Nederlandse leergangen voor begrijpend lezen. Hij constateert dat er in deze leergangen in het algemeen geen sprake is van een systematische aanpak van het begrijpend lezen. Een min of meer uitgewerkte systematische didactiek van het begrijpend lezen wordt volgens hem alleen gevonden in de leergangen "De leesmarkt" en "Met inzicht leren lezen". In beide leergangen wordt volgens Daandels eigenlijk maar één vaardigheid in een stap-voor-stap-procedure uitgewerkt nl. het maken van samenvattingen.

Een werkgroep onder leiding van Van den Berg en Te Lintelo (1974) heeft aan de hand van drie doelstellingenschema's de teksten en de daarbij behorende opdrachten van vijf leesmethoden geanalyseerd. Per methode worden de frequenties aangegeven waarmee de verschillende doelstellingen voor het begrijpend en studerend lezen in de opdrachten of verwerkingsopgaven voorkomen (vgl. Hatcher, 1971). Bovendien wordt van elke tekst en de daarbij behorende opdrachten een aantal kenmerken beschreven zoals leesbaarheid, omvang, inhoud en dergelijke. Opgemerkt moet worden dat deze analyse sterk constate-rend van aard is.

Zoals reeds eerder vermeld, analyseerden wij in een eerder onderzoek twee Nederlandse leesprogramma's nl. "Correct leren lezen" (Van der Schoot et al., 1973) en "Met inzicht leren lezen" (Van Eeuwijk-Rijnen et al., 1973). In "Correct leren lezen" wordt een poging gedaan om vooral de technische aspecten van het lezen systematisch te ontwikkelen en te verbeteren op basis van de modellen van Gray (1960) en Goodman (1970). Het programma "Met inzicht leren lezen" is hoofdzakelijk gericht op drie aspecten van het model van Gray nl. begrijpend lezen, kritisch lezen en assimilerend lezen en gebaseerd op de

taxonomie van Barrett (vgl. Clymer, 1968). Beide programma's zijn min of meer systematisch opgebouwd en gebaseerd op modellen die nog voor een groot deel empirisch moeten worden getoetst. Voor meer informatie verwijzen we naar Deelrapport 2 (Aarnoutse et al., 1976).

Uit het voorgaande zou men kunnen afleiden dat de ontwikkeling van programma's voor begrijpend lezen in Nederland nog grotendeels is gebaseerd op voorwetenschappelijke kennis en ervaring en dat over het effect van programma's nog niets te melden valt. Deze conclusie is niet juist wanneer men enkele programma's of leergangen in aanmerking neemt die momenteel ontwikkeld worden. We denken hier onder andere aan de constructie van een leergang voor begrijpend lezen binnen het project Onderwijs en Sociaal Milieu te Rotterdam (Nuy, 1980) en aan de ontwikkeling van een leergang voor het begrijpend lezen van zakelijke teksten door medewerkers van het Psychologisch Laboratorium te Utrecht (Bol, 1979; 1980). Aan beide leergangen zullen we in het kort aandacht besteden.

De eerst genoemde leergang vormt een onderdeel van een omvattend taal-leesprogramma bestemd voor leerjaar 2 tot en met 6. De ontwikkelaars van deze leergang vatten lezen op als een cognitief en communicatief proces dat op verschillende niveaus van volledigheid kan functioneren (Nuy, 1980). Zij onderscheiden vijf niveaus van lezen nl. technisch lezen, begrijpend lezen, gericht lezen, kritisch lezen en creatief lezen. Verondersteld wordt dat elk hoger niveau vollediger is dan het voorgaande niveau. Aan begrijpend lezen onderscheiden de ontwikkelaars in totaal 16 vaardigheden en attitudes, die in de vorm van doelen worden geformuleerd. Bijna alle vaardigheidsdoelen worden geklassificeerd volgens het Structure-of-Intellect model van Guilford en passen volgens Nuy (1980) goed binnen de theorie van Kintsch en Van Dijk. Nuy voegt hier echter direct aan toe dat er vooralsnog geen duidelijkheid bestaat omtrent de vraag in hoeverre de verschillende vaardigheidsdoelen empirisch onderscheidbaar zijn, met andere woorden verschillende cognitieve taken en processen veronderstellen.

De leergang van het tweede leerjaar is gericht op drie vaardigheidsdoelen. Deze globale doelen zijn vertaald in een aantal specifieke leerdoelen, die cognitieve activiteiten beschrijven. Elk specifiek doel is verder geoperationaliseerd in een bepaald type taak (een korte tekst en een bepaald vraagtype). Verondersteld wordt dat de typen taken elk een itemdomein speci-

ficeren van in essentie vergelijkbare items.

Nuy (1980) vermeldt een aantal leerpsychologische en didactische elementen die aan de constructie van de leergang ten grondslag liggen. Een belangrijk leerpsychologisch aspect is volgens hem dat het leerproces wordt gericht op de uitvoering van specifieke taken. Structurering vormt een voorwaarde voor een optimale oriëntering op de taak (vgl. Otto, 1977). Belangrijke didactische elementen of leercondities zijn onder andere oriëntering op de taak, actualiseren van voorkennis, herhaling, feedback, lezen door activiteit en beloning.

Wat betreft de opbouw van de leergang in niveaus en cycli, in verschillende soorten lessen en toetsen vermelden we hier slechts dat deze zeer systematisch is. Bij de evaluatie van de leergang voor het tweede leerjaar is zowel aandacht besteed aan de input-, proces- en produktevaluatie. Voor de laatst genoemde vorm van evaluatie is helaas geen beginmeting verricht en ook geen vergelijkende meting. Dit betekent in feite dat over een mogelijk effect van de leergang geen duidelijke uitspraken kunnen worden gedaan. Op grond van de beschikbare evaluatiegegevens concludeert Nuy dat er geen sprake is van een duidelijk leereffect als gevolg van het doorlopen van de leergang. Het gebrek aan eenduidigheid hierover wordt volgens hem veroorzaakt door een te geringe betrouwbaarheid van twee van de drie ontwikkelde toetsen.

Uit het voorgaande zal duidelijk zijn dat de leergang die door Nuy c.s. wordt ontwikkeld voor een groot deel is gebaseerd op de "skill-approach". De leergang die door Bol et al. (1979; 1980) wordt ontwikkeld, is op andere uitgangspunten gebaseerd. Zakelijke of informatieve teksten zijn volgens Bol teksten waarin de overdracht van kennis omtrent de realiteit een centrale plaats inneemt. Een schrijver analyseert een wereld op basis van mentale constructen (concepten, relaties, klassen, systemen en dergelijke) en reconstrueert deze in verbale beschrijvingen overeenkomstig zijn methode van onderzoek van die wereld (observatie, manipulatie, redenering en dergelijke). Een lezer moet over dezelfde mentale constructen en methoden van onderzoek en beschrijving beschikken om de uitspraken van de schrijver te begrijpen.

Volgens Bol (1980) zijn slechts weinig leerlingen in de hogere leerjaren zich bewust van de sociale achtergrond van zakelijke teksten. Veel leerlingen hebben moeite om geschreven teksten te plaatsen in het kader van de zo-

genaamde communicatieve driehoek (schrijver-tekst-lezer). Het belangrijkste probleem is echter dat de leerlingen de methoden van onderzoek en beschrijving die auteurs van zakelijke teksten hanteren niet beheersen.

Op basis van het cultuur-historische denken van Vygotsky en de onderwijs-leertheorie van Gal'perin komt Bol tot de volgende opzet van een programma voor het leren lezen van zakelijke teksten. Eerst worden de leerlingen bewust gemaakt van de sociale achtergrond van geschreven zakelijke teksten. Door gemeenschappelijke analyses van communicatiesituaties en door reflectie op hun eigen communicatieve handelingen ontdekken de leerlingen de algemene componenten en relaties binnen communicatiesituaties. Vervolgens leren de leerlingen de methoden van onderzoek en beschrijving die auteurs van zakelijke teksten hanteren. De eerste stap bestaat uit de analyse van de kenmerken van objecten. De leerlingen onderzoeken concrete objecten en beschrijven de kenmerken daarvan in een schema (de leerling als onderzoeker). Daarna wordt het schema in een tekst omgezet (de leerling als schrijver). Tenslotte moeten de leerlingen een tekst waarin alleen beschrijvingen van kenmerken van objecten voorkomen, terugbrengen tot een schema van kenmerken (de leerling als lezer). De tweede stap betreft de samenstelling van objecten, waarbij dezelfde procedure wordt toegepast als bij de eerste stap. De volgende stappen hebben betrekking op de vergelijking van objecten, de klassificatie van objecten, veranderingen van objecten etc. Nadat op deze wijze teksten met feitelijke beschrijvingen aan de orde zijn gekomen, worden de teksten uitgebreid met verklarende beschrijvingen en communicatieve toevoegingen. De essentie van het programma is "that the pupils learn to operate with written information in a communicative setting on the basis of firstly performing research activities themselves, then describing the results of their inquiries, and finally reading informative texts in which the same methods of inquiry and description are used as they learned to apply themselves" (Bol, 1980, pag. 9).

Het programma zelf bestaat uit een aantal blokken of modellen, die drie tot zes maanden in beslag nemen bij twee lessen per week. Elk blok bestaat uit een docentenboek, een leerlingenboek en een toetsenboek. Ieder blok wordt voorafgegaan door een pre-test en gevolgd door een post-test.

Bol (1979; 1980) heeft onderzoek gedaan naar het effect van een aantal blokken in enkele derde, vierde en vijfde klassen van twee experimentele

scholen en twee controlescholen. Uit de gegevens blijkt dat bij twee blokken een significant effect werd gevonden. Dit resultaat is hoopgevend al moet het met de nodige voorzichtigheid worden gehanteerd gezien de vrij lage betrouwbaarheid van de ontwikkelde paralleltests.

Het overzicht van programma's voor begrijpend lezen in de Verenigde Staten en Nederland is te beperkt om generaliserende uitspraken te doen. De volgende conclusies hebben dan ook alleen betrekking op de bestudeerde programma's en onderzoeken. Deze conclusies zijn: (a) het merendeel van de programma's voor begrijpend lezen is gebaseerd op voorwetenschappelijke kennis en ervaring, (b) bij de ontwikkeling van de programma's wordt slechts in beperkte mate gebruik gemaakt van formatieve en summatieve evaluatie, (c) het effect van de programma's is niet groot en vaak moeilijk interpreteerbaar.

3.1.3. Het effect van strategieën en experimentele programma's

In een uitvoerig overzicht van onderzoek op het gebied van begrijpend leesonderwijs onderscheiden Tierney en Cunningham (1980) twee soorten onderzoek. Het eerste soort onderzoek heeft betrekking op de vraag hoe het lezen van teksten bevorderd kan worden met andere woorden welke interventies, activiteiten of strategieën geschikt zijn om het begrijpen, onthouden en integreren van specifiek schriftelijke informatie bij leerlingen te verbeteren. Het tweede soort onderzoek heeft betrekking op de meer fundamentele vraag in hoeverre het mogelijk is de algemene en specifieke bekwaamheden van leerlingen op het gebied van begrijpend lezen te bevorderen, zó dat deze bekwaamheden ook bij het lezen van nieuwe teksten worden aangewend.

Wat betreft het eerste soort onderzoek maken Tierney en Cunningham onderscheid tussen studies die betrekking hebben op interventies die plaatsvinden vóór het lezen van een tekst, tijdens het lezen van een tekst en na het lezen van een tekst. Interventies of strategieën om het begrijpen van een tekst te verbeteren voordat de tekst wordt gelezen, zijn onder andere het uitbreiden van achtergrondkennis door de introductie van nieuwe begrippen, het activeren van achtergrondkennis en het richten van de aandacht door middel van "advance organizers", het formuleren van leerdoelen en vragen vooraf, het formuleren van vragen en doelen door de leerlingen en toevoegingen zoals tekeningen en titels. Interventies of strategieën om het begrijpen van een tekst tijdens het lezen te verbeteren zijn onder andere het oproepen van

voorstellingen of beelden omtrent mogelijke gebeurtenissen in een verhaal, het invoegen van vragen in een tekst, zelf vragen stellen, luidop lezen en het gebruik van procedures à la Stauffer en Manzo. Interventies of strategieën om het begrijpen van een tekst na het lezen te verbeteren, zijn onder andere het stellen van vragen naar aanleiding van de tekst, het geven van feedback en het voeren van discussies in een groep of klas. Volgens Tierney en Cunningham kan men uit de resultaten van een groot aantal onderzoeken naar het effect van strategieën om het begrijpen van een tekst te verbeteren slechts een zeer algemene conclusie trekken nl. dat toepassing van deze strategieën over het algemeen meer effect heeft dan wanneer dat niet gebeurt. Zij voegen hieraan toe dat "whatever positive effects practices or strategies tend to have, in general, their effects are likely to be fragile across such variables as texts, teachers, and students" (1980, pag. 41).

Het tweede soort onderzoek heeft betrekking op de vraag of het mogelijk is om bij leerlingen kennis, vaardigheden en strategieën te ontwikkelen, die ook bij het lezen van nieuwe teksten (wordt) worden aangewend. Deze vraag ligt ook aan het onderhavige onderzoek ten grondslag. Wat betreft dit soort onderzoek constateren Tierney en Cunningham dat er wel onderzoek is verricht naar het effect van leesprogramma's maar dat het onmogelijk is om vast te stellen "if the instructional components of these programs were responsible for whatever differences were observed, due to the variability across programs with respect to materials, grouping patterns, inservice training for teachers, and parent involvement" (1980, par. 44). In plaats van allerlei evaluatiestudies te vermelden, bespreken Tierney en Cunningham enkele onderzoeken met betrekking tot "metacomprehension", inferentie-training en het leren van tekststructuren. Deze onderzoeken geven volgens hen aan op welke wijze onderzoek op het gebied van "teaching reading comprehension" in de toekomst moet worden uitgevoerd.

In verband met het onderhavige onderzoek bespreken we in het volgende drie onderzoeken die Tierney en Cunningham vermelden met betrekking tot inferentie-training, een onderzoek naar het effect van een studievoordigheidsprogramma en enkele onderzoeken op het gebied van het leren van logische inferenties.

Hansen (1980; 1981) deed een onderzoek naar de effectiviteit van twee interventietechnieken bedoeld om de inferentiële leesbekwaamheid van kinderen

van het tweede leerjaar te verhogen. Uitgaande van de veronderstelling dat kinderen bij het lezen van schriftelijke informatie vaak te weinig gebruik maken van hun kennis en ervaring en daardoor niet spontaan tot inferenties komen, ontwikkelde Hansen twee interventiemethoden. De strategiemethode was er op gericht om vóór het lezen van de tekst zoveel mogelijk verbindingen tot stand te brengen tussen de voorkennis en de tekstuele informatie in de hoop dat kinderen dan spontaan inferenties zouden maken. De vraagmethode had tot doel de spontaneïteit van het infereren bij kinderen te bevorderen door veel oefening te geven in het beantwoorden van inferentiële vragen.

De 24 proefpersonen werden in drie groepen verdeeld; één groep kreeg instructie volgens de strategiemethode, een andere groep kreeg instructie volgens de vraagmethode en een derde groep, die als controlegroep fungeerde, kreeg onderwijs volgens de procedure van de "basal reader". Gedurende een periode van 10 weken kreeg elke groep 10 verhalen aangeboden. Elke groep kreeg eerst een introductie in nieuwe begrippen (les 1), daarna voorbereidende activiteiten op het te lezen verhaal (les 2), vervolgens het eigenlijke verhaal met vragen (les 3) en tenslotte oefeningen in decoderen (les 4). Als meetinstrumenten werd een pre-test, een test na elk verhaal, een post-test en de Stanford Achievement Test voor lezen afgenomen.

Wat betreft de scores op de test na elk verhaal bleek dat de twee experimentele groepen significant hoger scoorden dan de controlegroep en dat de groep met de strategiemethode significant hoger scoorde dan de groep met de vraagmethode. De post-test en de Stanford leestest waren bedoeld om het transfer-effect van de training te meten bij teksten die nieuw waren voor de kinderen. Wat betreft de scores op deze test werd slechts een gering effect geconstateerd ten gunste van de experimentele groepen, met name bij de groep met de vraagmethode.

Als tweede onderzoek vermelden Tierney en Cunningham de studie van Gordon (1979). In dit onderzoek, waarbij 42 leerlingen van het vijfde leerjaar waren betrokken, werd het effect van twee interventiemethoden vergeleken, die beide gericht waren op het verbeteren van de bekwaamheid van de leerlingen om gebruik te maken van voorkennis en tekstuele aanwijzingen. Eén interventiemethode was gericht op de ontwikkeling van voorkennis en kennis van tekststructuren. De tweede interventiemethode had tot doel de leerlingen strategieën te leren voor het maken van inferenties. Uit de resultaten bleek vol-

gens Tierney en Cunningham dat de tweede interventiemethode het meeste effect had met name op de transfer-taken.

Het derde onderzoek betreft een studie van Day (1980), die naging wat het effect is van een training in het samenvatten van teksten waarbij wel of geen expliciete aanwijzingen werden gegeven. Vier groepen studenten kregen elk een andere instructie aangeboden. Eén groep werd aangemoedigd om samenvattingen te maken en de hoofdgedachten uit teksten te halen; een tweede groep kreeg instructie in een aantal regels; een derde groep kreeg dezelfde instructie als de tweede groep maar werd daarbij aangemoedigd; een vierde groep kreeg instructie in een aantal regels en in de toepassing daarvan. Uit dit onderzoek bleek dat instructie in een aantal regels voor het maken van samenvattingen de bekwaamheid van de studenten het sterkst beïnvloedde en dat het effect van een dergelijke training varieerde met de begaafdheid van de studenten. Volgens Day hebben minder begaafde studenten een meer expliciete training nodig.

Naar aanleiding van de drie bovengenoemde onderzoeken met betrekking tot inferentie-training concluderen Tierney en Cunningham dat het mogelijk is om bij leerlingen kennis, vaardigheden en strategieën te ontwikkelen, die ook bij het lezen van nieuwe teksten (wordt) worden gebruikt. Het aantal studies waarin dit transfer-effect wordt aangetoond, is echter nog bijzonder gering.

Dat het niet eenvoudig is het proces van het zelfstandig verwerven en verwerken van schriftelijke informatie te verbeteren, blijkt uit een onderzoek van Kok, Boonman en Beukhof (1980). Op basis van de propositiebenadering van Kintsch en Van Dijk (1978) en de schemabenadering van Rumelhart et al. (1978) ontwikkelden deze onderzoekers een studievoordigheidsprogramma (STUBA), dat tot doel heeft het studeergedrag van leerlingen van het vijfde en zesde leerjaar te verbeteren. Dit programma bestaat uit een leergang studievoordigheden, een introductieprogramma, en een studeerwijzer (Beukhof et al., 1979). De leergang studievoordigheden omvat 14 lessen en heeft tot doel bij de leerlingen die begrippen en vaardigheden te ontwikkelen, die noodzakelijk zijn om de studeerwijzer te kunnen hanteren. Het introductieprogramma bestaat uit 5 lessen en heeft tot doel de studeerwijzer en het gebruik ervan bij de leerlingen te introduceren. De studeerwijzer zelf bestaat uit een aantal richtlijnen die de leerlingen bij het bestuderen van een tekst moeten

volgen en die tezamen een heuristische procedure vormen. Deze richtlijnen hebben tot doel de relevante inhoudelijke voorkennis te mobiliseren en de leerlingen te helpen bij het organiseren en reduceren van een tekst tot een coherente semantische macrostructuur.

Eén van de belangrijkste vraagstellingen van het onderzoek van Kok, Boonman en Beukhof (1980) was in hoeverre het aanbieden van een heuristische procedure voor tekstverwerking het opnemen en onthouden van informatie uit een tekst bevordert. Het onderzoek werd uitgevoerd bij leerlingen van het vijfde leerjaar. De proefpersonen van één van de experimentele condities kregen een training van 4 maanden waarna zij een tekst kregen aangeboden waarbij gebruik werd gemaakt van Studeerwijzer I uit het studievoordigheidsprogramma. Uit de gegevens bleek dat de leerlingen die een tekst met behulp van deze studeerwijzer analyseerden significant minder informatie hadden opgenomen dan de controlegroep. Uit de retentiemetingen bleek verder dat alle proefpersonen de opgenomen informatie in gelijke mate hadden onthouden.

In verband met het onderhavige onderzoek ligt het voor de hand dat nu een aantal onderzoeken wordt besproken naar het effect van trainingen of programma's met betrekking tot het afleiden van de hoofdgedachte uit een tekst, het afleiden van de bedoeling van de hoofdpersoon en het trekken van conclusies uit logische redeneringen. Dit is echter slechts in beperkte mate mogelijk, daar dit soort onderzoek nog zeer schaars is. Wat betreft het identificeren en afleiden van de hoofdgedachte zijn verschillende onderzoeken uitgevoerd, waarbij werd nagegaan wat leerlingen van de lagere school ten aanzien van deze taken presteren (vgl. Otto & Barrett, 1968; Van Blaricom & White, 1975 en 1976; Brown & Smiley, 1977). Voor zover bekend bestaat er echter nog geen experimenteel onderzoek naar het effect van programma's op dit gebied. Wat betreft het afleiden van de bedoeling van de hoofdpersoon in een tekst hebben we geen enkel onderzoek op het niveau van het lager onderwijs kunnen vinden. Dit geldt niet voor het afleiden van logisch geldige informatie. Zoals in hoofdstuk II is vermeld, hebben Kohnstamm (1934), Hill (1961), Ennis en Paulus (1965), Roberge (1970), Carroll (1975) en Ennis (1976) onderzoek uitgevoerd naar de bekwaamheid van kinderen om conclusies uit logische redeneringen te trekken. Op grond van deze onderzoeken hebben we vrij nauwkeurig kunnen vaststellen welke redeneervormen bij kinderen van het vierde leerjaar voor toetsing en training in aanmerking kunnen komen. Behalve deze onderzoe-

ken bestaan er op dit gebied ook enkele onderzoeken naar het effect van trainingen of programma's.

Eén van de meest bekende en omvangrijke studies is het onderzoek van Wolf, King en Huck (1968), dat onder meer tot doel had na te gaan of kritische leesvaardigheden bij kinderen van de lagere school kunnen worden ontwikkeld. Uitgaande van de veronderstelling dat kritisch lezen een aspect vormt van begrijpend lezen en van een operationele definitie van kritisch lezen in de vorm van een lijst van vaardigheden, ontwikkelden de onderzoekers zes experimentele leereenheden en een test voor kritisch lezen (de Critical Reading Test) bestemd voor leerjaar 1 tot en met 6. Een van de leereenheden had betrekking op redeneren (Logic) en was gericht op de ontwikkeling van twee vaardigheden nl. het herkennen en evalueren van de geldigheid van geschreven informatie en het herkennen en evalueren van de betrouwbaarheid van geschreven informatie. Uit het onderzoeksverslag van Wolf, King en Huck et al. (1967) blijkt dat in deze leereenheid onder andere de geldige redeneervormen *modus ponens* en *modus tollens* en de ongeldige redeneervormen *conversie* en *inversie* aan de orde komen. De items van de Critical Reading Test werden in drie secties gegroepeerd waaronder de sectie "Logic". De items van deze sectie "tested the subject's ability to detect fallacies and propaganda techniques used in printed materials and his ability to evaluate the internal consistency of an argument..." (Wolf, King & Huck, 1968, pag. 463).

Het onderzoek werd uitgevoerd bij leerlingen van 24 klassen van 7 lagere scholen. De klassen waren gelijkelijk verdeeld over leerjaar 1 tot en met 6 en over de experimentele en controleconditie. De leerkrachten van de experimentele klassen gaven, nadat ze eerst zelf instructie hadden ontvangen, twee lessen per week in kritisch lezen gedurende een geheel schooljaar. De leerkrachten van de controleklassen gaven, eveneens na instructie, in dezelfde tijd les in kinderliteratuur. Uit het onderzoek bleek dat de scores van de experimentele groep in alle leerjaren significant hoger waren op de Critical Reading Test dan die van de controlegroep. Uit een verdere analyse bleek dat bij de experimentele groep alleen de scores op de items van de sectie "Logic" voor elk leerjaar significant hoger waren. Een ander gegeven was dat er geen significant verschil tussen de experimentele en controlegroep werd gevonden wat betreft de scores op een norm-referenced leestest. Op grond van

de resultaten van dit onderzoek concluderen Wolf, King en Huck dat kinderen van leerjaar 1 tot en met 6 kritisch kunnen leren lezen met name wat betreft "logical reasoning".

Lowerre en Scandura (1973-1974) ontwikkelden een systematische procedure voor de constructie van diagnostische tests en onderwijsmaterialen met betrekking tot logische principes ("logic inference rules"). Vijf dimensies werden geïdentificeerd om de moeilijkheidsgraad van een bepaalde tekst te specificeren. Voor elk logisch principe kon de moeilijkheidsgraad van teksten worden geschat waarbij de lezer al of niet de correcte logische inferentie kon maken. Met behulp van deze procedure ontwikkelden de onderzoekers onderwijsmaterialen en tests voor leerlingen op het leesniveau van het derde en vierde leerjaar van de lagere school. De betreffende materialen en tests bevatten twee logische principes. Eén van deze principes had betrekking op de geldige redeneervorm *modus ponens* en de ongeldige redeneervorm *conversie*. De materialen varieerden over twee dimensies nl. complexiteit van de context en aanwezigheid van de premissen in de tekst. Het onderzoek werd uitgevoerd bij 40 leerlingen van het tweede, derde en vierde leerjaar, die gelijkelijk waren verdeeld over de experimentele en controleconditie. De experimentele groep kreeg onderwijs via een aantal werkboeken en een cassetteband. Uit het onderzoek bleek dat de scores van de experimentele groep significant hoger waren dan die van de controlegroep. Dit gold vooral voor de items die betrekking hadden op de redeneervormen *modus ponens* en *conversie*. Uit het onderzoek bleek dat de assumpties die aan de ontwikkelde procedure ten grondslag lagen, werden ondersteund.

Uit beide onderzoeken kan men afleiden dat kinderen van de lagere school met succes bepaalde logische principes met behulp van systematisch ontwikkelde programma's kunnen leren.

3.1.4. Opzet van de experimentele programma's

Zoals reeds eerder vermeld, worden in dit onderzoek drie experimentele programma's ontwikkeld op basis van de "skill-approach", hetgeen betekent dat elk programma is gericht op het leren beheersen van één bepaalde leestaak én de deeltaken die aan deze (eind)taak kunnen worden onderscheiden. In de programma's zal niet alleen aandacht worden besteed aan de ontwikkeling van gestructureerde kennis maar ook en vooral aan de ontwikkeling van str-

tegische of procedurele kennis, d.w.z. kennis met betrekking tot de wijze waarop de betreffende leestaak en haar deeltaken kan respectievelijk kunnen worden uitgevoerd. In afwijking van het Wisconsin Design zullen de betreffende programma's worden ontwikkeld voor klassikaal gebruik.

Voor de opzet of het ontwerp van deze programma's wordt in eerste instantie gebruik gemaakt van de resultaten van de literatuurstudie, die in verband met de specificatie van de drie leestaken (het afleiden van de hoofdgedachte uit een informatieve tekst, het afleiden van de bedoeling van de hoofdpersoon uit een fictieve tekst, en het trekken van conclusies uit logische redeneringen) is uitgevoerd. Zoals we gezien hebben, resulteerde deze studie in een vrij nauwkeurige omschrijving van deze taken en van de belangrijkste componenten of aspecten daarvan. Welke cognitieve processen bij de uitvoering van deze taken een rol spelen, kon echter niet of in onvoldoende mate worden omschreven. Voor het ontwerp van de drie programma's zal verder zo veel mogelijk gebruik worden gemaakt van gegevens van onderzoek en aanwijzingen van ontwikkelingswerk op het gebied van de betreffende taken. Zoals uit het voorafgaande hoofdstuk en de voorgaande paragraaf blijkt, zijn op het gebied van hoofdgedachte en redeneren verschillende onderzoeken uitgevoerd en is ook enig ontwikkelingswerk verricht. Voor de opzet van het programma Hoofdgedachte zullen we vooral gebruik maken van de onderzoeksgegevens van Otto en Barrett (1968), van Van Blaricom en White (1975; 1976) én van de ervaringen zoals die zijn neergelegd in verschillende programma's op het gebied van hoofdgedachte, met name van het Wisconsin Design (vgl. Otto, 1977). Voor de opzet van het programma Redeneren zullen we gebruik maken van de onderzoeksgegevens van onder anderen Ennis en Paulus (1965), Roberge (1970), Carroll (1975) en Ennis (1976) en in beperkte mate van het "reasoning" programma van het Wisconsin Design. Zoals uit het voorgaande blijkt kan bij de opzet van het programma Bedoeling geen gebruik worden gemaakt van onderzoeksgegevens en van ontwikkeld materiaal op dit gebied.

Wat betreft de opbouw van de drie programma's of leereenheden zullen we in grote lijnen de procedure volgen die Gagné (1974; 1977b) heeft voorgesteld om de leerstructuur van een te leren taak (i.c. leereenheid) vast te stellen. Via deze procedure, aangeduid met de term leertaakanalyse of leeranalyse, kunnen volgens Gagné de interne condities van een te leren taak worden geïdentificeerd en gearrangeerd. De leeranalyse omvat verschillende stappen die

hier in het kort worden vermeld.

De eerste stap bestaat uit een concrete beschrijving van de taak die de leerlingen op het einde van een bepaald programma of leereenheid moeten kunnen uitvoeren. Deze taakbeschrijving geeft het doel of de doelen weer die op het einde van een programma bereikt moeten worden. Een dergelijke taakbeschrijving "describes the performances which a program of learning is designed to insure that the learner acquires as a result of completing the program" (Gagné, 1977b, pag. 264). Zoals we gezien hebben, wordt in hoofdstuk II een beschrijving gegeven van de taken die de leerlingen op het einde van de programma's moeten kunnen uitvoeren. Elke taak wordt geformuleerd in termen van gedrag en inhoud.

De tweede stap bestaat hierin dat de omschreven taak wordt geplaatst binnen één van de vijf categorieën van leerresultaten of bekwaamheden die Gagné (1977b) onderscheidt nl. intellectuele vaardigheden, cognitieve strategieën, verbale informatie, motorische vaardigheden en attituden. Wat betreft de intellectuele vaardigheden onderscheidt Gagné de volgende subcategorieën: het leren van onderscheidingen, van concrete begrippen, van gedefinieerde begrippen, van regels en van samengestelde regels. Deze subcategorieën vertonen volgens hem een hiërarchische structuur. De drie leestaken waarvoor in dit onderzoek programma's en tests worden ontwikkeld, kunnen worden gecategoriseerd als intellectuele vaardigheden op het niveau van het leren toepassen van (complexe) regels.

Tijdens de derde stap wordt de eigenlijke leeranalyse uitgevoerd, waarbij de vereisten ("prerequisites") voor het leren van de gehele taak en elk van haar deeltaken worden geïdentificeerd. Gagné omschrijft een "prerequisite" als een "capability of prior learning which is incorporated into new learning" (1977b, pag. 268). Volgens hem zijn de duidelijkste voorbeelden van "prerequisites" te vinden binnen het domein van de intellectuele vaardigheden. Een complexe intellectuele vaardigheid kan worden geanalyseerd in ondergeschikte (subordinate) vaardigheden die bestaan uit meer eenvoudige regels en concepten. Zulke ondergeschikte vaardigheden zijn echte "prerequisites" d.w.z. ze zijn geïncorporeerd in de nieuw te leren vaardigheid die complexer is. "Prerequisites" kunnen worden opgevat als doelstellingen van de deeltaken waaruit een leertaak is samengesteld. Ze stellen de leerling in staat om de andere meer complexe doelstellingen te bereiken. Gagné spreekt

van "enabling objectives" (1977b, pag. 270).

De leeranalyse zoals hierboven beschreven, resulteert in een leerhiërarchie die, afhankelijk van haar functie en de aard van het object, meer of minder gedetailleerd kan zijn (vgl. Resnick & Beck, 1976). Volgens Gagné is een leerhiërarchie vooral van belang bij de "planning of sequences of instruction" (1977b, pag. 279). Bij de opzet van de drie programma's zal een leeranalyse worden uitgevoerd waarbij alleen de belangrijkste deeltaken of doelstellingen die de leerlingen achtereenvolgens moeten leren beheersen of bereiken, worden geïdentificeerd. Bij de identificatie en rangschikking van deze deeltaken ("enabling objectives") zal, zoals eerder vermeld, zoveel mogelijk rekening worden gehouden met gegevens van onderzoek en aanwijzingen van ontwikkelingswerk op het gebied van de betreffende taken.

Op basis van zijn model van informatieverwerking onderscheidt Gagné acht fasen binnen het leerproces en parallel hieraan acht externe condities of didactische maatregelen ("instructional events"), die van invloed kunnen zijn op de deelprocessen die tijdens de leerfasen plaatsvinden. Deze externe condities zijn: activeren van de motivatie, informeren van de leerling over het leerdoel, richten van de aandacht, stimuleren van de herinnering, begeleiding van het leren, versterken van de retentie, bevorderen van de overdracht van het geleerde, en uitlokken van de prestatie en verschaffen van feedback. Bij de verdere concretisering van de programma's in de vorm van leseenheden zal getracht worden deze externe condities zo goed mogelijk te realiseren. Met behulp van de try-outs van de programma's zal - zij het globaal - worden nagegaan in hoeverre de didactische maatregelen van invloed zijn op het leerproces.

3.2. Vaardigheden op het gebied van begrijpend lezen

Zoals reeds eerder vermeld, heeft dit onderzoek onder andere tot doel de realiteitswaarde vast te stellen van de onderscheiding van drie inferentiele leestaken. De vraag in hoeverre de betreffende taken empirisch onderscheidbaar zijn, wordt in dit onderzoek primair gesteld vanuit een didactische achtergrond nl. de legitimiteit van taakgericht onderwijs op het gebied van begrijpend lezen. Onderzoek naar de realiteitswaarde van de onderscheiding van leestaken op tekstniveau vertoont overeenkomst met eerder verricht

onderzoek, dat tot doel had na te gaan welke vaardigheden aan begrijpend lezen kunnen worden onderscheiden. In par. 3.2.1. beschrijven we in het kort welke vaardigheden een aantal auteurs binnen begrijpend lezen onderscheiden. Daarna volgt in par. 3.2.2. een overzicht van het onderzoek naar vaardigheden op het gebied van begrijpend lezen. Op het einde van deze paragraaf gaan we in op het verschil tussen dit soort onderzoek, dat sterk psychometrisch georiënteerd is, en het onderhavige onderzoek.

3.2.1. Vaardigheden van begrijpend lezen

Robinson (1966) onderscheidt drie typen modellen met betrekking tot lezen nl. (a) procesmodellen, waarin het verloop van het leesproces wordt weergegeven, (b) onderwijsmodellen, waarin procedures worden beschreven om het lezen te leren en te onderwijzen, (c) modellen van vaardigheden en bekwaamheden vereist voor de verschillende aspecten van lezen.

Het model van Gray (1960), dat later door Robinson (1966) werd uitgebreid, is in wezen een model van leesvaardigheden (vgl. Clymer, 1968). Volgens Gray zijn de activiteiten die bij het lezen plaatsvinden onder te brengen in een viertal categorieën nl. (a) woordperceptie, waarbij wordt ingesloten de uitspraak en het begrijpen van de betekenis van het woord, (b) begrijpen van de betekenis, (c) reactie op en evaluatie van de gedachten van de schrijver, (d) assimilatie van het gelezene ofwel de integratie van het gelezene met reeds aanwezige kennis. Bij elk van de vier hoofdaspecten van lezen somt Gray een aantal vaardigheden op die geleerd moeten worden. Hij onderscheidt drie niveaus van begrijpen nl. het begrijpen van de letterlijke betekenis van een tekst, het begrijpen van de impliciete betekenis van een tekst en het begrijpen van het belang en de implicaties van de ideeën van een schrijver. Voor elk niveau van begrijpend lezen somt Gray een aantal vaardigheden op zoals het volgen van en anticiperen op de gedachten van de schrijver, het bepalen van het doel van de schrijver en dergelijke. Robinson (1966) onderscheidt slechts twee niveaus van begrijpend lezen nl. het begrijpen van de letterlijke betekenis en de impliciete betekenis. Opvallend is dat Gray (1960), Robinson (1966) en Clymer (1968) geen duidelijke omschrijving geven van het begrip vaardigheid (skill). De vaardigheden die deze auteurs als voorbeeld noemen, worden over het algemeen in globale gedragstermen aangeduid. De waarde van het model van Gray - Robinson is vooral gelegen in de

ordering van belangrijke aspecten of componenten van lezen.

Barrett (1968) heeft een taxonomie voor het begrijpend lezen ontwikkeld. In dit klassificatieschema wordt begrijpend lezen in vijf categorieën van hoofdvaardigheden verdeeld. Deze vijf categorieën zijn (a) het begrijpen van de letterlijke betekenis (literal comprehension), (b) het herstructureren van een tekst (reorganization), (c) het begrijpen van een tekst via afleiding (inferential comprehension), (d) het beoordelen van een tekst (evaluation) en (e) het waarderen van een tekst (appreciation). Binnen elke hoofdcategorie onderscheidt Barrett een aantal taken in de vorm van doelen voor het lesonderwijs. De vijf hoofdcategorieën zijn geordend van gemakkelijk naar moeilijk. Ook de taken binnen elke hoofdcategorie zijn tot op zekere hoogte geordend van gemakkelijk naar moeilijk. Binnen de vaardigheid "inferential comprehension" onderscheidt Barrett de volgende taken: het afleiden van ondersteunende details, het afleiden van hoofdgedachten, het afleiden van de volgorde, het maken van vergelijkingen, het afleiden van oorzaak - gevolg relaties, het afleiden van karaktertrekken, het voorspellen van de afloop, en het interpreteren van figuurlijk gebruikte taal. Voor de volledige taxonomie verwijzen we naar Deelrapport 1 (Aarnoutse et al., 1975).

Het systeem van Barrett berust voor een groot deel op voorwetenschappelijke kennis en is empirisch nog niet gevalideerd. In ieder geval worden in vergelijking met het model van Gray - Robinson de hoofd- en deelvaardigheden nauwkeuriger omschreven. Een probleem bij dit systeem (en dat van Bloom, 1956) is de veronderstelling dat de (hoofd)vaardigheden een formele identiteit bezitten onafhankelijk van de aard, de inhoud en de structuur van het object i.c. de tekst. Men kan zich afvragen of het verantwoord is vaardigheden te onderscheiden en te ordenen zonder de kenmerken van het object waarop deze vaardigheden betrekking hebben i.c. de teksten hierbij te betrekken en te omschrijven. Een vaardigheid als het herkennen van de hoofdgedachte blijft vaag zolang niet duidelijk is wat de kenmerken zijn van het object i.c. de teksten waarvan de hoofdgedachte moet worden geïdentificeerd. Een klassificatieschema van taken op het gebied van begrijpend lezen dient ons inziens niet alleen een beschrijving te geven van de uit te voeren gedragingen maar ook van de teksten waarop deze gedragingen betrekking hebben.

In een poging het begrip "reading literacy" nader te analyseren en te definiëren, onderscheidt Bormuth (1973-1974) zeven categorieën of klassen van leesgedragingen. Deze categorieën omvatten gedragingen die betrekking hebben

op (a) decoderen, (b) letterlijk begrijpen, (c) het maken van inferenties, (d) kritisch lezen, (e) esthetische waardering, (f) leesflexibiliteit en (g) studerend lezen. De tweede categorie nl. het letterlijk begrijpen van een tekst omvat volgens Bormuth de volgende vier gedragingen: vocabulary meaning behaviors, sentence comprehension skills, anaphora comprehension behaviors en discourse comprehension behaviors (Bormuth, 1973-1974, pag. 20-21). Uit de context blijkt dat Bormuth de termen gedragingen en vaardigheden als synoniemen gebruikt en hiermee geen processen wil aanduiden die daaraan ten grondslag liggen. Opvallend is verder dat wat betreft de vaardigheden of gedragingen voor begrijpend lezen er een duidelijke overeenkomst bestaat tussen de categorieën van Bormuth en Barrett.

Zoals reeds eerder vermeld, is het Wisconsin Design for Reading Skill Development gebaseerd op de "skill-approach". Otto (1977) beschouwt de vaardigheden en het onderwijzen van vaardigheden "as the substance and the means for sharpening the process of developing reading comprehension" (pag. 196). Evenals Samuels (1976) beklemtoont hij dat de vaardigheden voor begrijpend lezen slechts middel zijn en geen doel.

In het Wisconsin Design wordt een onderscheid gemaakt tussen vaardigheden die een beroep doen op het convergente denken en vaardigheden die divergent denken veronderstellen. Het convergente denken leidt tot één oplossing van een probleem, terwijl bij het divergente denken meerdere oplossingen voor een bepaald probleem mogelijk zijn. Volgens Chester (1974) en Otto (1977) zijn beide typen van denken belangrijk en noodzakelijk in een totaal leesprogramma. De vaardigheden op het gebied van "comprehension" hebben betrekking op het convergente denken, terwijl de vaardigheden die divergent denken veronderstellen in het Wisconsin Design zijn gerangschikt onder de categorieën "interpretatief lezen" en "creatief lezen".

In het Wisconsin Design worden de vaardigheden voor begrijpend lezen niet zoals Gray (1960) en anderen voorstellen verdeeld in vaardigheden voor het letterlijk begrijpen en vaardigheden voor het afleiden van informatie. Chester et al. (1974) zijn van mening dat "the difference between inferring a main idea and inferring an outcome appears to be as great as the difference between determining the answer to a question when it involves factual recall and determining the answer to a question when it must be inferred. The difference between one inferential skill and another may at times be greater

than the difference between factual skills and inferential skills. Therefore, rather than having an inference skill in the Comprehension (element), we chose to develop skills which necessitate inferring the main idea, inferring an outcome, inferring an effect given a cause, inferring the meaning of a word given context clues..." (pag. 8). De opvatting van Chester sluit aan bij de verwachting die we eerder in hoofdstuk II formuleerden betreffende het mogelijk verschil tussen de geselecteerde leestaken.

In de "developmental edition" van het Wisconsin Design worden de vaardigheden op het gebied van begrijpend lezen in zes componenten verdeeld. Op grond van theoretische inzichten en empirisch onderzoek (vgl. Otto, 1977) worden sedert 1975-1976 vier componenten voor begrijpend lezen onderscheiden nl. woordbetekenis, zinsbetekenis, tekstbetekenis en sequentie. De component "tekstbetekenis" wordt verdeeld in twee deelcomponenten nl. "central thought" en "relationships/conclusions". De vaardigheden met betrekking tot "central thought" zijn gericht op de bekwaamheid om vast te stellen waar een tekst over gaat. De eerste drie vaardigheden (niveaus A-C) hebben betrekking op het identificeren van het algemene onderwerp (topic) van een tekst. De vierde vaardigheid (niveau D) is gericht op het identificeren van details die relevant en irrelevant zijn met betrekking tot de centrale gedachte. De vijfde en zesde vaardigheid (niveau E en F) hebben betrekking op het identificeren en afleiden van de hoofdgedachte. De doelstelling voor de laatste vaardigheid wordt als volgt geformuleerd: "The child identifies a given main idea statement of a passage that has no organizer and has both relevant and irrelevant details" (Van Blaricom, 1975, pag. 17). De laatste vaardigheid (niveau G) heeft betrekking op het genereren van de hoofdgedachte. De vaardigheden voor "relationships/conclusions" hebben betrekking op de bekwaamheid om expliciete informatie te identificeren en te begrijpen en op de bekwaamheid om deze informatie toe te passen bij de vorming van conclusies. De eerste drie vaardigheden (niveaus A-C) veronderstellen de synthese van gegeven informatie. Drie typen van relaties worden geleerd nl. descriptieve relaties, overeenkomst - verschil relaties en oorzaak - gevolg relaties. De volgende vier vaardigheden (niveaus D-G) hebben betrekking op het vormen van conclusies op basis van de gegeven informatie. Opvallend is dat in vergelijking met Gray (1960) en anderen de vaardigheden en doelen in het Wisconsin Design nauwkeuriger worden omschreven. De aard en de structuur van de tek-

sten worden expliciet vermeld (vgl. Van Blaricom & White, 1975).

3.2.2. Onderzoek naar vaardigheden van begrijpend lezen

In de jaren '30 en '40 werden in de Verenigde Staten verschillende onderzoeken op het gebied van begrijpend lezen uitgevoerd, waarbij voor het eerst factoranalyse werd toegepast. Deze studies hadden vooral tot doel na te gaan welke mentale vaardigheden of bekwaamheden aan begrijpend lezen kunnen worden onderscheiden.

Een belangrijk onderzoek uit die tijd was het onderzoek van Davis (1944). Via een literatuurstudie inventariseerde hij de belangrijkste vaardigheden van begrijpend lezen. Het resultaat was een lijst van enkele honderden vaardigheden. Na zuivering en ordening stelde Davis een ontwerp samen van "important skills of comprehension among mature readers" (Davis, 1972, pag. 662). Uit dit ontwerp selecteerde hij negen vaardigheden die hij beschouwde als "...basic to comprehension in reading" (Davis, 1944, pag. 186). Voor deze gepostuleerde vaardigheden ontwikkelde Davis tests met vijf keuze-items, die bij 421 eerstejaarsstudenten werden afgenomen. Het aantal items varieerde per test van 5 tot 60. De betrouwbaarheidscoëfficiënten van de tests varieerden aanzienlijk, nl. van .17 tot .90. Een ander zwak punt was dat verscheidene items, bedoeld om verschillende vaardigheden te meten, gebaseerd waren op eenzelfde leestekst. Na principale componentenanalyse concludeerde Davis dat minstens vijf factoren konden worden onderscheiden. Deze vijf factoren waren: kennis van de betekenis van de woorden, verbaal kunnen redeneren (het zien van relaties), gevolgtrekkingen kunnen maken, de structuur van een tekst kunnen volgen en het herkennen van de literaire technieken van een schrijver. Opvallend was dat de eerste twee factoren 89% van de totale variantie verklaarden. De eerste factor (kennis van de betekenis van woorden) verklaarde 80% van de variantie en de tweede factor (verbaal kunnen redeneren) 9% van de variantie. Op basis van dit onderzoek kwam Davis tot de conclusie dat begrijpend lezen niet één mentale vaardigheid is, maar dat aan begrijpend lezen minstens vijf afzonderlijke bekwaamheden (abilities) kunnen worden onderscheiden.

De resultaten van het onderzoek van Davis werden door Thurstone (1946) sterk in twijfel getrokken. Deze voerde een heranalyse uit op de data van Davis door een andere methode (de centroid methode) te gebruiken. Thurstone

kwam tot de conclusie dat de intercorrelaties van de testresultaten alleen verklaard konden worden "by a single common factor (reading ability) with remarkably small residuals" (Thurstone, 1946, pag. 188). Op grond hiervan is het volgens Thurstone zeer de vraag "whether any one would be tempted to extract a second factor from residuals so small as these - to say nothing of pulling out nine factors here" (pag. 187).

Ook andere onderzoekers vonden één algemene factor. Harris (1948) die in zijn onderzoek zeven vaardigheden van begrijpend lezen postuleerde, vond slechts één factor. Derrick (1953) vond na factoranalyse met de centroeide methode ook één factor, welke 44% van de variantie verklaarde. Bovendien vond Derrick geen verschil tussen vragen die betrekking hadden op feiten (factual questions), inferenties (inferential questions) en kritische oordelen.

De bovengenoemde factoranalytische studies geven geen empirische ondersteuning van de vraag of begrijpend lezen een samenhangend geheel van vaardigheden is of een enkelvoudige mentale vaardigheid met niet te onderscheiden deelvaardigheden. Deze onderzoeken zijn niet met elkaar te vergelijken en hebben nogal wat beperkingen zoals het gebruik van verschillende methoden van factoranalyse die leiden tot tegenstrijdige conclusies, de gebrekkige operationalisering van de vaardigheden, het gebrek aan validiteit van verschillende meetinstrumenten, de lage betrouwbaarheidscoëfficiënten van verschillende tests, het verschillend aantal items voor een test, het gebrek aan regels voor de ontwikkeling van de items, het verschillend aantal proefpersonen, de steekproef etc. (vgl. Chapman, 1971; Davis, 1972; Wardrop, 1976).

In 1968 werd door Davis een groot onderzoek uitgevoerd dat tot doel had aan te tonen dat begrijpend lezen uit verschillende mentale vaardigheden bestaat. Na de bestudering van empirisch onderzoek met betrekking tot de analyse van begrijpend lezen, selecteerde Davis acht vaardigheden die volgens hem fundamenteel zijn voor begrijpend lezen. Davis koos de volgende vaardigheden:

1. Recalling word meanings.
2. Drawing inferences about the meaning of a word from context.
3. Finding answers to questions answered explicitly or merely in paraphrase in the content.

4. Weaving together ideas in the content.
5. Drawing inferences from the content.
6. Recognizing a writer's purpose, attitude, tone and mood.
7. Identifying a writer's techniques.
8. Following the structure of a passage. (Davis, 1968, pag. 517).

Een vergelijking met de lijst van 1944 maakt duidelijk dat twee vaardigheden uit die lijst zijn gecombineerd tot vaardigheid 3 (Finding answers to questions answered explicitly or merely in paraphrase in the content).

Voor elke vaardigheid werden in totaal 40 items ontwikkeld of geselecteerd. Om het probleem van experimentele afhankelijkheid tussen antwoorden op items te vermijden, werd elk item gebaseerd op één afzonderlijke tekst. Voor elke vaardigheid werden twee paralleltests (Form A en B) samengesteld. In een experimentele try-out werden deze tests (in totaal 320 items) in twee sessies bij ongeveer 400 proefpersonen van de high school (twelfth grade) afgenomen. Van de 40 items voor elke vaardigheid werden vervolgens 24 items geselecteerd, die een hogere gemiddelde correlatie hadden met de totaalscore op die vaardigheid dan met de totaalscores op de andere zeven vaardigheden. Voor elke vaardigheid werden twee paralleltests (Form C en D) van elk 12 items samengesteld. Deze paralleltests werden met een interval van een of twee dagen bij ongeveer 1100 proefpersonen van de high school (twelfth grade) afgenomen. Elke proefpersoon maakte dus alle tests van Form C en D (in totaal 192 items). De analyse van de data was gebaseerd op 988 proefpersonen.

Voor de bepaling van het percentage unieke variantie van elk van de acht gepostuleerde vaardigheden maakte Davis (1968) gebruik van een "uniqueness analysis" techniek. De schattingen van de voor elke vaardigheid unieke variantie werden gemaakt door het kwadraat van de multiële correlatie van de ene variabele met alle overige variabelen van één af te trekken, nadat deze multiële correlatie met behulp van crossvalidatie (splitsing van de steekproef van proefpersonen, twee afname-momenten) was gecorrigeerd. De percentages unieke variantie voor elke vaardigheid staan vermeld in tabel 3.1.

Uit tabel 3.1. blijkt dat vaardigheid 1 het hoogste percentage unieke variantie heeft. Ongeveer 32% van de variantie van deze vaardigheid komt niet voor in een van de andere zeven vaardigheden. Vaardigheid 5 vertoont eveneens een hoog percentage unieke variantie. Ongeveer 20% van de variantie

Tabel 3.1. Overzicht van de percentages unieke variantie binnen de non-error variantie van acht vaardigheden.

Vaardigheid	Crossvalidatie over	
	items en dag	items
1. Recalling word meanings	35	29
2. Drawing inferences about the meaning of a word from context	- 1	8
3. Finding answers to questions answered explicitly or merely in paraphrase	13	7
4. Weaving together ideas in the content	5	5
5. Drawing inferences from content	23	18
6. Recognizing a writer's purpose, attitude, tone and mood	14	8
7. Identifying a writer's techniques	8	3
8. Following the structure of a passage	15	12

is uniek. De drie andere vaardigheden die nog behoorlijke percentages unieke variantie bezitten, zijn vaardigheid 8, 6 en 3.

Op basis van deze gegevens concludeert Davis (1968) dat begrijpend lezen bij volwassen lezers niet een enkelvoudige mentale vaardigheid is. De gegevens laten er volgens hem geen twijfel over bestaan "that substantial parts of the mental abilities used in eight skills judged to be of importance in comprehension are independent of one another" (Davis, 1968, pag. 542). Volgens Davis dient het voortgezet onderwijs de leerlingen vertrouwd te maken met de betekenis van zoveel mogelijk woorden (vaardigheid 1) en ze te stimuleren inferenties te maken uit datgene wat ze lezen (vaardigheid 5). Ook dienen er aparte oefeningen gegeven te worden voor de ontwikkeling van de vaardigheden 3, 6 en 8. Volgens Davis heeft het weinig zin om de vaardigheden met een geringe unieke variantie (2, 4 en 7) apart te oefenen.

Davis (1971; 1972) heeft op dezelfde data ook een principale componenten-analyse uitgevoerd. Na varimax rotatie vond hij vijf factoren. De eerste twee factoren waren: kennis van woordbetekenissen (vaardigheid 1) en het maken van inferenties uit de tekstinhoud (vaardigheid 5).

De data van het onderzoek van Davis zijn door verschillende onderzoekers

opnieuw geanalyseerd met andere analyse-technieken. Thorndike (1971; 1973-1974) analyseerde de correlatiematrices opnieuw met behulp van principale componentenanalyse met betrouwbaarheidscoëfficiënten op de diagonaal in plaats van enen op de diagonaal. Uit deze analyse bleek dat de betreffende correlatiematrix volledig verklaard kon worden door drie factoren, waarbij de eerste factor 93% van de variantie verklaarde. Evenals bij Davis bleek ook hier dat de test voor woordbetekenis duidelijk verschillend was van de andere tests. Thorndike interpreteerde de eerste factor niet als kennis van woorden zoals waarschijnlijk Davis zou hebben gedaan, maar als "reasoning in reading". Volgens Thorndike is de woordenschattest nl. niet vrij van "reasoning" (denken). Hij verwijst hiervoor naar de woordenschat-subtest van de Stanford-Binet die in vergelijking met de andere subtests het hoogste correleert met het IQ. Een scherpe scheiding tussen "word knowledge" en "reasoning in reading" is volgens Thorndike dan ook niet juist. Volgens hem is lezen voor een groot deel een kwestie van denken ("reasoning").

Spearritt (1972) analyseerde met behulp van de maximum likelyhood factor-analytische procedures van Jöreskog de correlatiematrices van Davis. Na toepassing van het juiste model vond hij dat de vaardigheden 1, 5, 6 en 8 als afzonderlijke vaardigheden waren te onderscheiden. Met uitzondering van vaardigheid 3, die statistisch niet kon worden aangetoond, is dit resultaat in overeenstemming met de bevindingen van Davis (1968) verkregen met behulp van zijn "uniqueness analysis". Uit het onderzoek van Spearritt bleek -evenals uit de analyse van Davis (1968) en Thorndike (1971) - dat woordenschat (vaardigheid 1) het duidelijkst kon worden onderscheiden en dat de andere drie vaardigheden zeer hoog met elkaar correleerden. Uit het feit dat de andere drie vaardigheden zeer hoog met elkaar correleren, trekt Spearritt de volgende conclusie: "Thus, although certain comprehension skills can be differentiated, present types of reading comprehension tests, as distinct from knowledge tests, largely measure one basic ability, which may well correspond to the label of "reasoning in reading"". (Spearritt, 1972, pag. 110). Deze conclusie komt voor een deel overeen met die van Thorndike (1971).

Chapman (1971) onderscheidt drie theorieën of modellen over vaardigheden op het gebied van begrijpend lezen nl. de theorie van de onafhankelijke vaardigheden, de theorie van de globale vaardigheid en de theorie van de hiërarchische vaardigheden. De theorie van de onafhankelijke vaardigheden

postuleert dat begrijpend lezen bestaat uit een aantal afzonderlijke vaardigheden, die onafhankelijk van elkaar en zonder een bepaalde volgorde kunnen worden onderwezen en geleerd. Chapman karakteriseert deze theorie als een multifactor theorie; begrijpend lezen bestaat uit verschillende factoren of vaardigheden die geen of weinig relaties met elkaar hebben. Volgens Chapman is Davis een vertegenwoordiger van de "independent skills theory". Davis (1972) spreekt dit echter tegen door er op te wijzen dat uit zijn onderzoek blijkt dat de vaardigheden nauw met elkaar samenhangen. De theorie van de globale vaardigheid postuleert dat begrijpend lezen één enkelvoudige algemene vaardigheid is, waaraan geen deelvaardigheden kunnen worden onderscheiden. Chapman typeert deze theorie als een unifactor theorie. Volgens deze theorie bestaan er hoge correlaties tussen de testcores voor begrijpend lezen; deze hoge correlaties kunnen slechts door één algemene factor worden verklaard. Zoals we gezien hebben, zijn Thurstone (1946), Harris (1948), Derrick (1953) en anderen van mening dat begrijpend lezen één algemene vaardigheid is. Davis (1972) verwerpt deze theorie door te wijzen op de resultaten van zijn onderzoek. De theorie van de hiërarchische vaardigheden postuleert dat begrijpend lezen uit een aantal vaardigheden bestaat, die verschillen in complexiteit en die aan elkaar gerelateerd zijn omdat de eenvoudige vaardigheden componenten zijn van de meer complexe. Verondersteld wordt dat vaardigheden voor begrijpend lezen in niveaus kunnen worden geordend volgens het principe van de complexiteit van het gedrag dat noodzakelijk is om elke vaardigheid te leren. Vaardigheden die sterk verschillen in complexiteit correleren volgens deze theorie niet hoog met elkaar. Wat het onderwijs betreft impliceert deze theorie dat het verwerven van een eenvoudige vaardigheid een noodzakelijke voorwaarde is voor de ontwikkeling van een meer complexe vaardigheid. Expliciet aandacht besteden aan de ontwikkeling van een complexe vaardigheid betekent volgens deze opvatting dat de meer eenvoudige vaardigheden tegelijkertijd worden geoefend. Volgens Barrett (1968), Chapman (1969) en anderen bestaat begrijpend lezen uit een aantal vaardigheden die hiërarchisch geordend zijn.

Volgens Chapman (1971) kunnen vaardigheden op het gebied van begrijpend lezen op twee manieren worden geconceptualiseerd nl. als mentale processen (bijvoorbeeld de bekwaamheid om de hoofdgedachte van een tekst te identificeren) en als reacties op taaleenheden. In navolging van Bormuth (1970) de-

finieert Chapman de vaardigheden van "literal reading comprehension" in termen van reacties op linguïstische eenheden nl. woorden, zinnen, anaphora, relaties tussen zinnen en inferenties. Uitgaande van deze eenheden onderscheidt Chapman vijf vaardigheden voor "literal reading comprehension" die volgens haar hiërarchisch zijn geordend. Deze vaardigheden zijn: vocabulary comprehension, sentence comprehension, anaphora comprehension, intersentence comprehension en inference comprehension (Chapman, 1971, pag. 71). In haar theoretisch model postuleert Chapman dat er een lineaire hiërarchische relatie bestaat tussen deze vijf vaardigheden. Dit betekent dat de eenvoudige vaardigheden zoals "vocabulary comprehension" en "sentence comprehension" noodzakelijke voorwaarden zijn voor het verwerven van complexe vaardigheden.

In een zorgvuldig opgezet onderzoek heeft Chapman (1971) geprobeerd het model van de "hierarchical skills theory" te toetsen. Dit onderzoek onderscheidt zich in twee opzichten van de vorige onderzoeken. In de eerste plaats is het onderzoek niet op volwassen lezers gericht, maar op leerlingen van de vijfde klas van de lagere school. In de tweede plaats is het onderzoek experimenteel van opzet: twee experimentele groepen kregen 10 weken lang elke dag gedurende 25 minuten een zelfinstructieprogramma bestaande uit 50 lessen. Eén experimentele groep kreeg onderwijs in "sentence comprehension", terwijl een andere experimentele groep instructie kreeg in "intersentence comprehension". Een controlegroep kreeg in dezelfde tijd een programma voor rekenen aangeboden. Een test bedoeld om de vijf bovengenoemde vaardigheden te meten, werd voor en na de instructie afgenomen. Uit dit onderzoek bleek dat de leerlingen die instructie hadden ontvangen in "sentence comprehension" en "intersentence comprehension" significant beter op de tests voor deze vaardigheden presteerden dan leerlingen van de controlegroep. Een transfer-effect van deze twee vaardigheden naar de drie andere vaardigheden, die niet waren onderwezen, werd niet gevonden. Verder bleek dat er sprake was van een hiërarchische relatie tussen de twee onderwezen vaardigheden. Onderwijs in "sentence comprehension" bleek een positieve invloed te hebben op "intersentence comprehension". Onderwijs in "intersentence comprehension" bleek echter geen positieve invloed te hebben op "sentence comprehension". Uit dit gegeven concludeert Chapman dat "acquisition of the simpler skill seems prerequisite to proficiency in the more complex, but not vice versa" (Chapman, 1971, pag. 182). Bovendien bleek dat de vaardigheden geen hiër-

chische structuur vertoonden wanneer deze onafhankelijk van elkaar over de verschillende teksten werden getoetst. Wanneer de volgorde van de vaardigheden binnen de teksten werd getoetst, bleek er echter sprake te zijn van "some evidence of a hierarchy" (Chapman, 1971, pag. 182).

Op grond van het onderzoek van Chapman kan men niet concluderen dat er een hiërarchische relatie bestaat tussen de verschillende vaardigheden voor begrijpend lezen. Het onderzoek bevat wel aanwijzingen in de richting van de "hierarchical skills theory". Het is echter de vraag of voor deze hiërarchie een lineair-additieve combinatie van vaardigheden, zoals Chapman voorstelt, de meest aangewezen is. Wellicht zal veeleer gedacht moeten worden aan partiële hiërarchieën. Ook het onderzoek van Andrich en Godfrey (1978-1979) geeft steun aan de theorie dat begrijpend lezen uit vaardigheden bestaat, die hiërarchisch zijn geordend. Andrich en Godfrey analyseerden de scores van 188 volwassen proefpersonen op de leestest van Davis (Form D) waarbij gebruik werd gemaakt van het latente trekmodel van Rasch. Uit deze analyse bleek dat 76 van de 96 items in overeenstemming waren met dit model. Uit dit gegeven concluderen de onderzoekers dat de betreffende test een één-dimensionele vaardigheid meet. Bij weglating van de items die kennis van woorden meten, bleek verder dat de overgebleven 67 items vier verschillende clusters op de ene dimensie vormden. Volgens Andrich en Godfrey kunnen deze clusters worden geïnterpreteerd als vaardigheden die verschillende niveaus vormen van een hiërarchische structuur. Opgemerkt moet worden dat de test van Davis (1968) niet is ontwikkeld met de bedoeling een verzameling van hiërarchische vaardigheden te meten. De resultaten van Andrich en Godfrey moeten dan ook als voorlopig worden beschouwd. Uit het onderzoek van Chapman (1971) en Andrich en Godfrey (1978-1979) zou men voorzichtig kunnen afleiden dat begrijpend lezen uit een aantal vaardigheden bestaat die nauw met elkaar samenhangen en waarvan een deel wellicht hiërarchisch geordend is.

Zoals uit het literatuuroverzicht van Davis (1972) en Chapman (1971) blijkt, is er veel onderzoek gedaan waarin met behulp van factoranalyse en regressie-analyse gezocht is naar de essentiële componenten of vaardigheden van begrijpend lezen. Met uitzondering van het onderzoek van Chapman zijn deze onderzoeken op een aantal punten bekritiseerd. In navolging van Raygor geeft Farr (1969) de meest fundamentele kritiek, die er op neerkomt, dat de resultaten van deze studies staan of vallen met de validiteit en betrouw-

baarheid van de meetinstrumenten die worden gebruikt. Hoewel in het onderzoek van Davis (1968) veel aandacht is besteed aan de ontwikkeling van betrouwbare instrumenten, rijzen er problemen bij de validiteit van deze tests. De oorzaak van deze problemen is gelegen in het feit dat de keuze van de vaardigheden voor begrijpend lezen voor een groot deel is gebaseerd op voorwetenschappelijke kennis en ervaring en niet of nauwelijks op grond van een theorie van begrijpend lezen. Het onderzoek van Davis (1968) en de analyse van Thorndike (1971) en anderen zijn niet alleen theoretisch onvoldoende onderbouwd, maar maken ook gebruik van analysemethoden, i.c. factoranalyse, die minder geschikt zijn (vgl. Andrich & Godfrey, 1978-1979). Volgens Samuels (1976) is het methodologisch zwakke punt van deze studies "the failure to differentiate between good and poor readers in the analysis. The fluent reader has mastered the subskills, combined them into higher units, so that the intercorrelations among subskills should be high, thus making reading seem to be but one skill-called reading" (pag. 173). Guthrie (1973) ontwikkelde speciale tests en analyseerde de intercorrelaties voor de goede en zwakke lezers afzonderlijk. Zoals voorspeld, vond Guthrie bij de goede lezers zeer hoge intercorrelaties die er op wijzen dat lezen één vaardigheid was geworden. Bij zwakke lezers werden lage intercorrelaties gevonden wat wijst op het bestaan van afzonderlijke vaardigheden. Tot slot moet worden opgemerkt dat de vaardigheden die Davis (1968) onderscheidt nogal algemeen van aard zijn. Verschillende vaardigheden worden niet duidelijk omschreven, terwijl een beschrijving van het object i.c. de tekst in het geheel ontbreekt. Wat Davis bijvoorbeeld precies bedoelt met de vaardigheid "Drawing inferences from content" is niet duidelijk. Bovendien veronderstelt hij dat een dergelijke algemene vaardigheid een formele identiteit heeft los en onafhankelijk van de aard, de inhoud en de structuur van de tekst.

Deze laatste opmerking geeft aanleiding om het verschil tussen het onderzoek van Davis (1968) en het onderhavige onderzoek nader te expliciteren. Uit hoofdstuk I blijkt dat dit onderzoek niet tot doel heeft na te gaan welke essentiële vaardigheden of bekwaamheden aan begrijpend lezen kunnen worden onderscheiden. Zoals reeds eerder vermeld, heeft het onderzoek onder andere tot doel de realiteitswaarde van de onderscheiding van drie inferentiële leestaken vast te stellen. Nagegaan wordt in hoeverre deze taken, die in de terminologie van Davis betrekking hebben op één vaardigheid nl. het maken

van inferenties uit een tekst, in werkelijkheid zijn te onderscheiden. Op grond van de lijst van vaardigheden van Davis zou men verwachten dat deze taken niet verschillen en via onderwijs een positieve invloed op elkaar hebben. Op basis van het idee dat er een sterk verband bestaat tussen het gedrags- en inhoudsaspect van een taak i.c. het afleiden en datgene wat moet worden afgeleid, verwachten we dat de drie geselecteerde taken verschillen en via onderwijs geen positieve invloed op elkaar uitoefenen. Zoals we in hoofdstuk II hebben gezien, verschillen de drie taken vermoedelijk ook wat betreft de gestructureerde kennis waarop tijdens de uitvoering een beroep kan worden gedaan.

Uit het voorgaande mag niet worden afgeleid dat de literatuurstudie omtrent vaardigheden op het gebied van begrijpend lezen geen invloed heeft gehad op het onderhavige onderzoek. Zoals uit het vervolg zal blijken, hebben we enkele ideeën overgenomen. In navolging van Davis (1968) hebben we bij de ontwikkeling van de taakgerichte tests elk item gebaseerd op één afzonderlijke tekst. In navolging van Chapman (1971) is voor een experimentele opzet van het onderzoek gekozen. Uit een recente publicatie van Rosenshine (1980) blijkt dat de problematiek van de onderscheiding van vaardigheden op het gebied van begrijpend lezen nog steeds actueel is. Het onderzoeksvoorstel van Rosenshine vertoont veel overeenkomst met de opzet van het onderhavige onderzoek.

3.3. Methoden en instrumenten voor het meten van begrijpend lezen

Zoals uit hoofdstuk I blijkt, heeft dit onderzoek onder andere tot doel de betrouwbaarheid en validiteit vast te stellen van vijf tests op het gebied van het begrijpend lezen van teksten. In par. 3.3.1. wordt aan de hand van de indeling van Carroll (1972) een beschrijving gegeven van de verschillende methoden voor het meten van begrijpend lezen. Daarna wordt in par. 3.3.2. aandacht besteed aan twee typen tests die in dit kader centraal staan nl. taakgerichte tests en cloze tests en aan onderzoek dat in verband met de validiteit van deze tests van belang is. Op basis van de bestudeerde onderzoeksliteratuur worden verwachtingen geformuleerd omtrent mogelijke verbanden tussen de te ontwikkelen tests en maten voor technisch lezen, begrijpend lezen, woordenschat en non-verbale intelligentie.

3.3.1. Methoden voor het meten van begrijpend lezen

Op basis van een inventarisatie van procedures die in psychometrisch en experimenteel onderzoek worden gehanteerd, heeft Carroll (1972) een classificatieschema ontworpen van methoden of procedures voor het meten van taalbegrip i.c. begrijpend lezen. In dit schema worden drie dimensies onderscheiden nl. de taken die de proefpersonen moeten uitvoeren, de typen metingen of observaties die worden verricht en de condities waaronder gemeten wordt. Carroll onderscheidt de volgende taken: (a) verslag doen over het al of niet begrepen hebben van de aangeboden informatie of over specifieke aspecten van de boodschap, (b) verslag doen over de juistheid van de aangeboden informatie of over de overeenkomst met een andere presentatie, (c) opdrachten uitvoeren door een niet verbale respons te geven, (d) ontbrekende elementen in de informatie aanvullen, (e) vragen beantwoorden op basis van de informatie, (f) herkennen van informatie of elementen daarvan bij een tweede aanbieding, (g) reproduceren van informatie in haar geheel of gedeeltelijk, in oorspronkelijke of gewijzigde vorm. Aan de hand van deze taken volgt nu een beschrijving van de verschillende methoden of procedures voor het meten van begrijpend lezen. Voor deze beschrijving hebben we naast het overzicht van Carroll ook gebruik gemaakt van publicaties van Bormuth (1970), Anderson (1972), Tuinman (1973-1974), Wesdorp (1978) en anderen.

Verslag doen over het al of niet begrepen hebben van de aangeboden informatie of over specifieke aspecten van de boodschap.

Subjectieve rapportages van proefpersonen kunnen volgens Carroll valide en nuttig zijn bij het meten van taalbegrip als er zekerheid bestaat over de "eerlijkheid" en de aandacht van de proefpersonen en als er tegelijkertijd metingen kunnen worden gedaan over bijvoorbeeld de beslissingstijd. Subjectieve verslaggeving kan gemakkelijk leiden tot onjuiste positieve resultaten wanneer de proefpersoon gelooft dat hij de informatie begrepen heeft, terwijl dat niet het geval is. De betrouwbaarheid van deze methode is gering met name bij kinderen. Volgens Carroll zijn de mogelijkheden van deze methode nog in onvoldoende mate onderzocht (vgl. Gibson & Levin, 1975). De methode kan volgens hem gebruikt worden om na te gaan welke elementen in een tekst het begrijpen moeilijk maken bijvoorbeeld bepaalde woorden, grammaticale constructies, zinsdelen. Door veranderingen in de tekst aan te brengen is het moge-

lijk subjectieve oordelen te relateren aan tekstenkenmerken zoals grammaticaaliteit, ambiguïteit, grammaticale complexiteit. Bij informele evaluatie-instrumenten zoals Reading Inventories wordt de methode van subjectieve verslaggeving soms toegepast (vgl. Farr, 1969).

Verslag doen over de juistheid van de aangeboden informatie of over de overeenkomst met een andere presentatie.

Volgens Carroll is de verificatie van informatie die gebaseerd is op een analyse van deze informatie of op een visuele voorstelling daarvan een zeer geschikte methode om het begrijpen te meten. Het juist beantwoorden van waar/onwaar vragen op uitspraken die letterlijk uit de tekst zijn overgenomen, is volgens Anderson (1972) nog geen bewijs voor het begrijpen van dergelijke uitspraken. Het juist reageren op geparafraseerde uitspraken is volgens hem wel een bewijs dat de oorspronkelijke uitspraken zijn begrepen. Gibson en Levin (1975) wijzen er op dat het formuleren van goede parafrasen niet gemakkelijk is. De methode van het verifiëren van informatie wordt vooral toegepast op zins- en woordniveau. In woordenschattests wordt van de proefpersoon vaak gevraagd het woord te selecteren, dat in betekenis overeenkomt met een gegeven woord (vgl. de Woordenschattest van Stijnen, 1975).

De verificatie van informatie met behulp van visuele voorstellingen is een methode, die onder andere is toegepast door Trabasso (1972) en Bransford en McCarrell (1974). Aan de proefpersonen wordt de vraag gesteld welke visuele voorstelling identiek is met de aangeboden verbale informatie. Het werken met visuele voorstellingen (picture verification procedure) heeft twee nadelen nl. (a) het maken van goede tekeningen is moeilijk en tijdrovend, (b) het visueel representeren van verbale informatie is slechts in beperkte mate mogelijk.

Opdrachten uitvoeren door een niet verbale respons te geven.

Tests waarbij proefpersonen schriftelijke opdrachten krijgen aangeboden en deze vervolgens moeten uitvoeren in de vorm van handelingen (tekeningen, werkzaamheden), kunnen volgens Carroll in hoge mate valide en betrouwbaar zijn. Een juiste uitvoering van de in de tekst beschreven opdracht vormt het bewijs dat de tekst is begrepen. Om de validiteit te verzekeren moet de opdracht van dien aard zijn, dat deze niet kan worden uitgevoerd tenzij de proefpersonen de instructie hebben begrepen. Bovendien moet de opdracht ge-

makkelijk zijn uit te voeren. Ook Farr (1970) en Wesdorp (1978) zijn van mening dat deze methode zeer valide metingen kan opleveren. Wesdorp (1978) geeft een overzicht van de diverse soorten opdracht-toetsen en wijst op de mogelijkheden en beperkingen van deze methode. Een beperking van deze methode is, dat ze slechts bij bepaalde verbale informatie kan worden gebruikt, en dat het geheugen de testresultaten kan beïnvloeden. De mogelijkheid is nl. niet uitgesloten dat de proefpersonen de instructie hebben begrepen maar deze zijn vergeten op het moment dat de opdracht moet worden uitgevoerd. Een goed voorbeeld van deze methode is de schoolvorderingentest "Schriftelijke Opdrachten" van Brus en Van Bergen (1973).

Ontbrekende elementen in de informatie aanvullen.

Het meest bekende voorbeeld van deze methode is de cloze procedure, die aanvankelijk vooral gebruikt werd om de leesbaarheid van teksten vast te stellen. De cloze procedure bestaat hierin dat in een tekst een aantal woorden volgens vooraf opgestelde regels wordt weggelaten bijvoorbeeld elk vijfde woord, elk zelfstandig naamwoord of functiewoord. Vervolgens krijgen de proefpersonen de tekst aangeboden met de opdracht de weggelaten woorden op te schrijven. Volgens Carroll is de cloze procedure algemeen aanvaard als een maat om iemands begrip van een bepaalde tekst vast te stellen, al staat de validiteit van deze procedure nog ter discussie. Zo zijn er aanwijzingen dat de cloze scores meer afhankelijk zijn van grammaticale dan van semantische aanwijzingen. Volgens Carroll is de cloze procedure in haar gewone vorm "too crude to permit measuring the degree to which the individual comprehends particular lexical or grammatical cues, or possesses a knowledge of specified linguistic rules. It probably depends to a considerable extent on inferential processes." (1972, pag. 19). In de volgende paragraaf zullen we aandacht besteden aan onder andere de validiteit van verschillende cloze tests.

Naast de normale cloze procedure bestaat ook de "progressive cloze" procedure waarbij de proefpersoon elk volgend woord in een tekst moet raden, nadat de tekst eerst gelezen is. De methode wordt gebruikt om de moeilijkheidsgraad van teksten vast te stellen. De methode om ontbrekende elementen in informatie aan te vullen wordt ook toegepast in tests, waarbij de proefpersonen zinnen moeten aanvullen of zinnen in een goede volgorde moeten plaatsen.

Vragen beantwoorden op basis van de informatie

Het beantwoorden van vragen is een methode die in het onderwijs veel wordt gebruikt om na te gaan of de informatie is begrepen. In de meeste gestandaardiseerde tests voor begrijpend lezen worden ook teksten met vragen aangeboden. Voor de beantwoording van de vragen blijven meestal de teksten beschikbaar (vgl. Aarnoutse et al., 1977).

Het formuleren van duidelijke vragen is een kwestie waaraan vaak veel te weinig aandacht wordt besteed. Dit geldt zowel voor het opstellen van open vragen als voor de constructie van items (vgl. Wesdorp, 1978). Uit onderzoek van onder anderen Tuinman (1973-1974) blijkt dat items vaak niet tekstafhankelijk zijn. Dit betekent dat de items door de proefpersonen kunnen worden beantwoord zonder dat de tekst is gelezen. Vaak doen items een beroep op kennis waarover de proefpersoon al beschikt of bevatten ze zoveel aanwijzingen dat het antwoord eigenlijk al gegeven is. Het komt ook voor dat het begrijpen van de vragen (items) moeilijker is dan de tekst zelf; bovendien zijn vragen (items) vaak niet onafhankelijk van elkaar.

Verschillende auteurs onderscheiden verschillende soorten van vragen. Van der Hoeve, Kohnstamm en Van Veen (1935) spreken van (a) oriënteringsvragen, (b) ordeningsvragen, (c) gebonden combinatievragen, (d) vrije combinatievragen en (e) vragen waarbij het antwoord als waarde-oordeel moet worden gegeven. Palland (1935) onderscheidt twee soorten vragen nl. reproducerende en producerende vragen.

Volgens Bormuth (1970) moeten vragen op een objectieve wijze d.w.z. volgens bepaalde expliciet geformuleerde regels worden gegenereerd. Bormuth toont aan dat eenvoudige vragen met behulp van de principes van de transformationele generatieve grammatica op een objectieve wijze kunnen worden ontwikkeld. Hij onderscheidt vier typen vragen, die hier worden geïllustreerd aan de hand van de zin /de jongen reed op het ros/. De vier typen vragen op het niveau van het begrijpen van zinnen zijn:

- a. de letterlijke vraag (rote question). Wie reed op het ros?
- b. de getransformeerde vraag (transform question). Door wie werd het ros bereden?
- c. de geparafraseerde vraag (semantic substitute question). Wie reed op het paard?
- d. de getransformeerde en geparafraseerde vraag (compound question). Door wie

werd het paard bereden?

Anderson (1972) onderscheidt in totaal zes typen vragen. De eerste vier typen komen overeen met de typen van Bormuth. Volgens Anderson zijn de eerste twee typen vragen van Bormuth niet geschikt om vast te stellen of iemand bepaalde informatie heeft begrepen. De twee typen vragen die Anderson verder onderscheidt, zijn bedoeld om het begrijpen van concepten en principes te meten. Het begrijpen van concepten kan volgens hem worden vastgesteld met behulp van vragen waarbij algemene termen worden vervangen door bijzondere termen. Het begrijpen van principes kan worden gemeten met behulp van vragen waarbij specifieke termen moeten worden vervangen door algemene termen. Bij de Hoofdgedachtetest wordt van de laatst genoemde procedure gebruik gemaakt.

Herkennen van informatie of elementen daarvan bij een tweede aanbieding

De herkenningmethode is een traditionele methode om het leren en onthouden te meten. Na presentatie van bepaalde verbale informatie krijgen de proefpersonen elementen van de oorspronkelijke informatie samen met nieuwe of veranderde elementen aangeboden met de opdracht die elementen te selecteren, die reeds eerder werden aangeboden. Carroll wijst er uitdrukkelijk op dat bij de herkenningmethode het onthouden een belangrijkere rol speelt dan het begrijpen. Carroll beschouwt de "chunked comprehension" test van Carver als een toepassing van de herkenningmethode. Carver (1970) laat leerlingen een tekst van vier à vijf alinea's lezen. Onmiddellijk daarop wordt een meerkeuze-test afgenomen zonder dat de tekst beschikbaar is. Elk item van de meerkeuze-test bestaat uit alternatieven die gevormd worden door de "chunks" (woordgroepen), die in de oorspronkelijke tekst voorkomen. Eén woordgroep is echter veranderd; de leerlingen moeten aanstrepen welk alternatief niet conform de oorspronkelijke tekst is. Carver heeft twee onderzoeken uitgevoerd om aan te tonen dat de "chunk comprehension" test begripend lezen meet. Volgens Farr (1970) en Carroll (1972) is deze test niet geschikt om begripend lezen te meten; de test meet volgens hen eerder het onthouden dan het begrijpen van verbale informatie. Ondanks veel kritiek is Rystrom (1972) van mening dat verder onderzoek naar de methode van Carver gewenst is.

Reproducere van informatie in haar geheel of gedeeltelijk, in oorspronkelijke of gewijzigde vorm

Zoals we in hoofdstuk II hebben gezien, krijgen proefpersonen in psycholinguïstisch onderzoek vaak de taak om de aangeboden informatie op een of andere wijze te reproducere. Afhankelijk van de aard van de taak en de condities waaronder die moet worden uitgevoerd, spelen in deze onderzoeken geheugenprocessen een rol en blijft het volgens Carroll moeilijk de processen van begrijpen en onthouden te isoleren.

Een variant van de reproductiemethode is het parafraseren, het reproducere van verbale informatie in eigen woorden. Deze methode is geschikt voor het meten van tekstbegrip voor zover de parafrasen op een objectieve manier kunnen worden gescoord. Parafraseren kan volgens Carroll ook een extra belasting betekenen voor de proefpersonen. Het maken van een samenvatting van een tekst kan worden opgevat als een vorm van parafrasering. Wesdorp (1978) besteedt de nodige aandacht aan de samenvatting, die in veel eindexamens van het voortgezet onderwijs als opdracht voorkomt.

3.3.2. Instrumenten voor het meten van begrijpend lezen

In het onderwijs worden over het algemeen twee soorten van instrumenten gebruikt voor het meten van begrijpend lezen nl. formele, gestandaardiseerde en informele, niet gestandaardiseerde instrumenten (vgl. Farr, 1969). Bij de formele instrumenten worden vooral twee methoden gehanteerd nl. het beantwoorden van vragen op basis van de gegeven informatie en het aanvullen van de ontbrekende elementen in de informatie. Bij de informele instrumenten zoals Informal Reading Inventories worden meerdere methoden gebruikt om de leesbevoegdheid van leerlingen vast te stellen. In het volgende bespreken we eerst twee typen formele instrumenten voor het meten van begrijpend lezen nl. norm-referenced tests en criterion-referenced tests. Daarna besteden we aandacht aan de twee typen tests die hier centraal staan en aan onderzoek dat in verband met de validiteit van deze tests van belang is. Tenslotte worden enkele verwachtingen geformuleerd over mogelijke verbanden met andere tests.

Een norm-referenced test (N.R.T.) is een test waarbij de score van een leerling of een klas wordt vergeleken en geïnterpreteerd aan de hand van de scores van een aselecte steekproef die de norm van de test verschaftte. Een N.R.T. wordt meestal ontwikkeld om de verschillen tussen leerlingen wat

betreft een algemene bekwaamheid of vaardigheid vast te stellen. Met behulp van een N.R.T. kunnen de leerlingen worden gerangschikt en kan aan de hand van de relatieve norm worden vastgesteld welke leerlingen tot de besten, de zwaksten en tot de middengroep behoren. De items van een N.R.T. moeten dan ook goed discrimineren tussen goede en zwakke leerlingen.

In de Verenigde Staten is een groot aantal norm-referenced tests voor begrijpend lezen ontwikkeld. Farr en Anastasiow (1971) geven een uitvoerige en kritische beschrijving van vijf bekende Amerikaanse tests voor begrijpend lezen. In Deelrapport 3 (Aarnoutse et al., 1977) hebben we vier Amerikaanse tests beschreven nl. de Gates MacGinitie Reading Tests, de Progressive Achievement Tests of Reading, de Stanford Achievement Test-Reading en de Sequential Tests of Educational Progress-Reading. In Nederland bestaan ook enkele norm-referenced tests voor begrijpend lezen en woordenschat zoals de test Schriftelijke Opdrachten van Brus en Van Bergen (1973), de Stilleestest van Bakker (1972) en de Woordenschattest van Stijnen (1975). Deze tests zijn voor Nederland geijkt.

Norm-referenced tests voor begrijpend lezen hebben een beperkte functie. Volgens Farr (1969) zijn ze betrouwbaar om leerlingen te vergelijken in termen van "general reading achievement" (pag. 85) en geven ze een indicatie voor eventueel verder diagnostisch onderzoek. Farr (1969; 1971) wijst er nadrukkelijk op dat de meeste norm-referenced tests voor begrijpend lezen niet voor diagnostische doeleinden gebruikt mogen worden; de subtests zijn hiervoor niet voldoende valide en betrouwbaar. Volgens Calfee en Drum (z.j.) zijn norm-referenced tests voor begrijpend lezen niet ontwikkeld met het doel de leerkracht te helpen bij het nemen van didactische beslissingen op korte termijn. Een norm-referenced test meet geen specifieke doelstellingen of taken op het gebied van begrijpend lezen; het is een test die voor een groot deel "curriculum-free" is (vgl. Dahl, 1971). Bormuth (1975) levert scherpe kritiek op de bestaande norm-referenced tests voor begrijpend lezen. Volgens hem zijn deze tests te veel gericht op hogere cognitieve processen (reasoning processes) en te weinig op het letterlijk begrijpen van informatie. Bovendien zijn de tests niet volgens objectieve, expliciet geformuleerde regels ontwikkeld en kunnen de ruwe scores moeilijk worden geïnterpreteerd vanwege het ontbreken van gelijke intervallen en een zinvol nulpunt. De tests zijn volgens Bormuth ook niet gevoelig genoeg om de groei in prestaties op korte termijn

vast te stellen en te rigide wat betreft de vormgeving. In hoeverre deze kritiek helemaal terecht is, gezien de beperkte kennis op het gebied van begrip en lezen, laten we hier buiten beschouwing. Een feit is in ieder geval dat norm-referenced tests weinig informatie verschaffen aan de leerkracht bij het arrangeren en evalueren van onderwijsleersituaties. Een test die deze informatie wel kan verschaffen, is een criterion-referenced test.

Een criterion-referenced test (C.R.T.) is een test, waarbij de score van een leerling of een klas rechtstreeks wordt geïnterpreteerd in termen van een absolute standaard of een absoluut criterium (vgl. Glaser & Nitko, 1971). Een C.R.T. heeft meestal betrekking op één doelstelling of een beperkt aantal doelstellingen die op korte termijn worden nagestreefd. Een C.R.T. is een instrument waarmee de leerkracht onder meer kan vaststellen welke leerlingen de doelstelling of doelstellingen hebben bereikt en welke niet. Als leerlingen 80% of meer van de items van een C.R.T. goed maken, neemt men meestal aan dat de doelstelling is bereikt en dat er sprake is van "mastery" (Bloom et al., 1971).

In de Verenigde Staten zijn in de zeventiger jaren verschillende evaluatieprogramma's voor het leesonderwijs ontwikkeld. Deze programma's of systemen bestaan uit een grote verzameling doelstellingen voor het leesonderwijs en items gebaseerd op deze doelen. Met behulp van deze doelen en items kunnen criterion-referenced tests worden ontwikkeld. In Deelrapport 2 (Aarnoutse et al., 1976) hebben we drie evaluatie-programma's uitvoerig beschreven nl. "Instructional Objectives Exchange" (I.O.X.) met de daarbij behorende tests, "Mastery: an evaluation tool" en het "System for Pupil and Program Evaluation and Development" (S.P.P.E.D.). Kenmerkend voor deze systemen is dat de keuze en de ordening van de doelstellingen theoretisch nog onvoldoende is gefundeerd en dat gegevens over de betrouwbaarheid en validiteit van de criterion-referenced tests grotendeels ontbreken. Hetzelfde kan gezegd worden van twee criterion-referenced tests die we in Deelrapport 3 (Aarnoutse et al., 1977) beschreven nl. "Prescriptive Reading Inventory" en "Diagnosis: an instructional aid". Een goede uitzondering vormen de criterion-referenced tests die binnen het Wisconsin Design for Reading Skill Development zijn ontwikkeld. Deze tests worden theoretisch in ieder geval meer verantwoord, terwijl ook onderzoek wordt gedaan naar de betrouwbaarheid van de tests. Elke test heeft betrekking op één doelstelling en bestaat uit twee parallelvormen, die als pre- en post-test worden gebruikt. Het aantal items per test varieert van

15 tot 20. Over het algemeen zijn de KR-20 betrouwbaarheidscoëfficiënten .80 of hoger (vgl. Aarnoutse et al., 1976).

Het zal duidelijk zijn dat criterion-referenced tests die voldoende valide en betrouwbaar zijn een belangrijke functie kunnen vervullen bij het arrangeren en evalueren van onderwijsleersituaties voor begrijpend lezen. Volgens Tuinman (1978) zullen de tests dan wel een organisch onderdeel moeten vormen van de onderwijssituatie en niet zoals in de Verenigde Staten vaak het geval is een "superimposed nicety" (pag. 168). Om te voorkomen dat er eindeloze doelstellingen en tests worden geformuleerd respectievelijk ontwikkeld, is het volgens hem noodzakelijk dat er onderzoek wordt gedaan naar de vraag welke doelstellingen een bijdrage leveren aan of voorwaarde zijn voor het bereiken van het algemene doel. Met andere woorden welke doelen "criticale doelen" zijn voor het algemene doel nl. goed lezen of volwassen leesgedrag. Tuinman (1978) twijfelt er trouwens aan of het praktisch mogelijk is om voor lezen echte criterion-referenced tests te ontwikkelen. Volgens hem is het praktisch niet mogelijk "to calibrate language or define relevant characteristics so that items of comparable difficulty and complexity can be generated from them" (pag. 171).

Met dit probleem hangt een ander probleem direct samen nl. de vaststelling van het criterium. Tuinman (1978) wijst er op dat het criterium meestal wordt vastgesteld op basis van conventie en intuïtie en niet op grond van empirisch onderzoek. Otto (1973) is van mening dat bepaalde doelen een hoge graad van beheersing vereisen in verband met het verdere leesproces, terwijl dit voor andere doelen niet het geval behoeft te zijn. Calfee en Drum (z.j.) vragen zich af waarom steeds een criterium van 80% wordt aangehouden.

Een ander probleem in verband met criterion-referenced tests is psychometrisch van aard en hangt samen met het feit dat de klassieke testtheorie voor een groot deel is gebaseerd op relatief meten waarbij gestreefd wordt naar een grote spreiding van de scores. Bij een C.R.T. wordt de score van een leerling rechtstreeks geïnterpreteerd in termen van een absoluut criterium en staat het verkrijgen van scorevariantie niet centraal. Wanneer het onderwijs inderdaad effectief is, dan zal de variantie van de scores bij een C.R.T. gering of zelfs nul zijn, hetgeen betekent dat de conventionele technieken om de betrouwbaarheid en validiteit van een C.R.T. vast te stellen niet kunnen worden toegepast en dat andere benaderingswijzen moeten worden ontwikkeld.

Volgens De Klerk (1979) komen in de praktijk echter zelden situaties voor waarbij een scorevariantie van nul bereikt wordt. Ook Terwel (1973) is van mening dat bij formatieve evaluatie voldoende scorevariantie overblijft om de conventionele technieken van de klassieke testtheorie te gebruiken.

Wat betreft de ontwikkeling van alternatieve psychometrische technieken in verband met "criterion-referenced measurement" en de problemen die hiermee samenhangen, verwijzen we kortheidshalve naar publicaties van Popham (1975; 1978) en Warries (1979). Opgemerkt moet worden dat een aantal auteurs zoals Wijnen (1973) en Ebel (1975) twijfelen aan de noodzaak van een alternatieve meettechnologie. Zij geven ook bij formatieve evaluatie de voorkeur aan de toepassing van de klassieke testtheorie. Een werkgroep onder leiding van Mommers (Bulte et al., 1979) toont bijvoorbeeld aan dat ook met de klassieke testtheorie een differentiële criteriumtoets ontwikkeld kan worden (in dit geval voor vermenigvuldigen).

Taakgerichte tests

Uit het bovenstaande kan reeds worden afgeleid dat in dit onderzoek geen criterion-referenced tests voor het begrijpend lezen van teksten worden ontwikkeld. Naast de problemen die Tuinman (1978) met betrekking tot de constructie van deze tests voor lezen signaleert, heeft ook de opvatting van onder anderen Calfee en Drum (z.j.) hierbij een rol gespeeld nl. dat het goed mogelijk is om specifieke op vaardigheden of taken gerichte leestests te ontwikkelen op basis van de principes van de klassieke testtheorie. Volgens hen kunnen met behulp van deze principes betrouwbare tests worden ontwikkeld, die een belangrijke functie kunnen vervullen bij het evalueren en arrangeren van onderwijsleersituaties voor begrijpend lezen. Eén probleem dat inherent is aan ook specifieke norm-referenced tests blijft echter onopgelost nl. de vraag in hoeverre dit type tests gevoelig genoeg is om de groei in prestaties op korte termijn vast te stellen. Deze vraag kan slechts via empirisch onderzoek worden beantwoord.

Zoals reeds eerder vermeld, worden in dit onderzoek drie tests ontwikkeld die betrekking hebben op drie verschillende inferentiële leestaken. Tests die gericht zijn op het meten van bepaalde, specifieke leestaken noemen we in dit verband taakgerichte tests. In een eerder onderzoek (Aarnoutse et al., 1977) duiden we dergelijke tests aan met de term doelstellinggerichte tests. Omdat

doelstellinggerichte tests vaak worden vereenzelvigd met criterion-referenced tests (vgl. Dahl, 1971) spreken we hier - ter voorkoming van misverstand - van taakgerichte tests.

Zoals Farr (1969) terecht heeft gesteld, dient bij de constructie van valide tests voor begrijpend lezen de keuze van de taken (vaardigheden) gebaseerd te zijn op een theorie van begrijpend lezen. In hoofdstuk II is nagegaan in hoeverre de keuze van inferentiële leestaken gebaseerd kon worden op een theorie over begrijpend lezen en inferentieprocessen. Gebleken is dat theorieën over begrijpend lezen nog niet voldoende zijn uitgewerkt en getoetst om op basis hiervan inferentiële leestaken te selecteren. Wel was het mogelijk een aantal inferentiële leestaken uit deze theorieën af te leiden en te formuleren. Zoals we gezien hebben, zijn uiteindelijk drie taken geselecteerd die passen binnen de beschreven theorieën en die voldoen aan de eerder gestelde criteria. Na de selectie van de taken werd vervolgens een literatuurstudie uitgevoerd met het doel de taken te specificeren en voor zover mogelijk theoretisch te funderen.

Zoals we gezien hebben, was het mogelijk de drie leestaken vrij nauwkeurig te specificeren wat betreft de kenmerken van de teksten. Voor de ontwikkeling van de Redeneertest konden specifieke regels worden opgesteld, terwijl voor de constructie van de Hoofdgedachtetest en de Bedoelingtest tamelijk nauwkeurige regels konden worden geformuleerd (vgl. Bormuth, 1970). Tengevolge van het ontbreken van een uitgewerkte theorie over begrijpend lezen en inferentieprocessen was het niet of nauwelijks mogelijk een beschrijving te geven van de processen die bij de uitvoering van de leestaken een rol spelen.

De procedure die tot dusver is gevolgd, had tot doel de validiteit van de te ontwikkelen tests zo veel mogelijk veilig te stellen. Zoals reeds eerder vermeld, zal in verband met de validiteit van de taakgerichte tests de samenhang worden nagegaan tussen deze tests en maten voor technisch lezen, begrijpend lezen, woordenschat en non-verbale intelligentie. Alvorens verwachtingen te formuleren, bespreken we enkele onderzoeken waarin de samenhang werd vastgesteld tussen technisch en begrijpend lezen, woordenschat en begrijpend lezen, en intelligentie en begrijpend lezen. We zullen ons hier uitdrukkelijk beperken tot slechts enkele voor ons doel relevante onderzoeken. Voor meer informatie verwijzen we naar publicaties van onder anderen Singer en Ruddell (1976), Gibson en Levin (1975), Guthrie (1977) en Calfee en Drum (1978).

Wat betreft de relatie tussen technisch lezen en begrijpend lezen is de laatste jaren veel onderzoek gedaan. Te denken valt aan onderzoek van onder anderen Golinkoff en Rosinsky (1976), Perfetti (1977), Perfetti en Lesgold (1977). Zoals gezegd beperken we ons hier tot slechts enkele voor ons doel relevante onderzoeken.

Brus en Voeten (1973) ontwikkelden de Eén-Minuut-Test om de technische leesvaardigheid te meten. Technisch lezen omschrijven zij als de vaardigheid in het ontsleutelen van gedrukte woorden. In verband met het onderzoek naar de validiteit van deze test werd voor de leerjaren 2 tot en met 6 ($n = \pm 700$ per leerjaar) onder andere de samenhang vastgesteld tussen de scores op de Eén-Minuut-Test en enkele tests voor begrijpend lezen. Als tests voor begrijpend lezen werd onder andere gebruik gemaakt van de Schriftelijke Opdrachten van Brus en Van Bergen (1973) en de Stilleestest van Bakker (1972). De correlatie tussen de Eén-Minuut-Test en Schriftelijke Opdrachten bedroeg in leerjaar 2, 3 en 4 respectievelijk .64, .46 en .42. De correlatie tussen de Eén-Minuut-Test en de Stilleestest bedroeg in leerjaar 3, 4 en 6 respectievelijk .50, .42 en .31. Uit deze gegevens blijkt dat er in het tweede leerjaar een aanzienlijke samenhang bestaat tussen technisch lezen en begrijpend lezen en dat deze samenhang in leerjaar 4 vrij laag is. Brus en Voeten (1973) concluderen uit deze en andere gegevens dat "met het voortschrijden van het basis-onderwijs de leesprestatie van de leerlingen zich differentieert. Het begrijpend lezen en het technisch lezen in de zin van de vaardigheid in het ontsleutelen van een tekst treden geleidelijk aan duidelijker naar voren als te onderscheiden aspecten van de leesvaardigheid" (pag. 44).

Zwarts (1977) voerde een onderzoek uit waarbij onder andere de samenhang werd nagegaan tussen technisch lezen en begrijpend lezen in het eerste leerjaar van de lagere school ($n = \pm 500$). Als maat voor technisch lezen gebruikte hij de Eén-Minuut-Test voor technische leesvaardigheid van Caesar (1970); voor begrijpend lezen werd gebruik gemaakt van de test Begrijpend Lezen I van Van Calcar, Tellegen en Soest (1970). Zwarts vond een vrij hoge correlatie (.69) tussen technisch en begrijpend lezen. In een vervolgonderzoek ($n = \pm 1600$) werd een toetsende factoranalyse uitgevoerd, waaruit bleek dat de beide constructen in het eerste leerjaar onderscheidbaar waren (vgl. Appelhof, 1979).

Danhof, Smelt en Van der Wissel (1978) voerden een onderzoek uit waarin de

samenhang werd nagegaan tussen technisch lezen en begrijpend lezen bij drie klassen van het derde leerjaar en drie klassen van het zesde leerjaar. Technisch lezen werd gemeten met behulp van de Eén-Minuut-Test van Brus van Voeten (1973) en de Leesvaardigheidstest van Wiegersma (1971); voor begrijpend lezen werd gebruik gemaakt van de Stilleestest van Bakker (1972). De correlaties tussen deze tests bedroegen in leerjaar 3 respectievelijk .62 en .61; in leerjaar 6 werden correlatiecoëfficiënten van respectievelijk .27 en .41 gevonden. Uit deze gegevens blijkt, evenals uit het onderzoek van Brus en Voeten (1973), dat technisch en begrijpend lezen zich in de loop van de lagere school in toenemende mate differentiëren.

Hoewel het begrijpen van de betekenis van woorden een belangrijk aspect vormt van de taalvaardigheid en leesvaardigheid, wordt er volgens Calfee en Drum (1978) relatief veel te weinig onderzoek gedaan naar "reading vocabulary" of de relatie tussen woordenschat en begrijpend lezen. Uit een onderzoek van Farr (1968) blijkt dat er een nauwe samenhang bestaat tussen subtests voor woordenschat en subtests voor begrijpend lezen. Er werden correlaties van .73 en .76 vastgesteld. Ook Pavlak (1973) constateert in zijn literatuuronderzoek dat zich over 24 jaar uitstrekt, dat "There was a substantive relationship between knowing multiple meanings of words and reading comprehension" (pag. 105). Volgens Thorndike (1971; 1973-1974) bestaat er een nauw verband tussen "word knowledge" en "reasoning in reading". Vellutino (1979) vermeldt een achttal studies waarin een substantiële samenhang werd gevonden tussen tests voor woordenschat en (technisch) lezen.

In het kader van het ijkingsonderzoek ontwikkelde Stijnen (1975) een test, die de lees-woordenschat (de receptief, schriftelijke woordenschat) van kinderen van het derde tot en met het zesde leerjaar van de lagere school meet. Stijnen berekende onder andere voor het vierde leerjaar de correlaties tussen deze woordenschattest en twee tests voor begrijpend lezen nl. de Schriftelijke Opdrachten van Brus en Van Bergen (1973) en de Stilleestest van Bakker (1972). Uit de gegevens blijkt dat er een aanzienlijke samenhang bestaat tussen de Woordenschattest en de Stilleestest (.65) en een hoge samenhang tussen de Woordenschattest en Schriftelijke Opdrachten (.76). Het feit dat er een aanzienlijke tot hoge samenhang wordt gevonden tussen de lees-woordenschat en begrijpend lezen wijst er op dat het begrijpen van de betekenis van woorden waarschijnlijk een zeer belangrijk aspect vormt van de leesvaardig-

heid. Kinderen die over een beperkte lees-woordenschat of algemener geformuleerd over een beperkt aantal georganiseerde kenniseenheden of schemata beschikken (vgl. Rumelhart en anderen), zullen naar verwachting veel moeite hebben met het begrijpend lezen van teksten.

Er is veel onderzoek gedaan naar de relatie tussen intelligentie en begrijpend lezen. Voor een overzicht van de relatie tussen intelligentie en technisch lezen in het eerste leerjaar van de lagere school verwijzen we naar een publicatie van Van Dongen en Van Leent (1981).

In een overzichtsartikel met als titel "IQ is and is not related to reading" toont Singer (1973) aan de hand van een aantal studies aan dat de relatie tussen intelligentie en aspecten van de leesvaardigheid in de loop der jaren verandert. Singer vermeldt een onderzoek van Durkin (1966) die in het eerste leerjaar een correlatie van .40 vond tussen intelligentie en lezen, terwijl in de daarop volgende jaren bij dezelfde groep leerlingen de correlatie toenam en in het vijfde leerjaar een waarde bereikte van .79. Een verklaring voor deze verandering van 16% naar 64% gemeenschappelijke variantie tussen intelligentie en lezen bestaat volgens Singer hierin "that, as reading achievement shifts from predominant emphasis on word recognition to stress on word meaning and comprehension, the mental functions being assessed by intelligence and reading tests have more in common" (pag. 23). Ook Farr (1969) wijst er in zijn overzicht op dat de samenhang tussen intelligentie en lezen toeneemt naarmate de leerlingen ouder worden en het begrijpend lezen een belangrijke rol gaat spelen. Zowel Singer (1973) als Farr (1969) vermelden verder een aantal onderzoeken waaruit blijkt dat de scores op tests voor begrijpend lezen hoger correleren met scores op verbale intelligentietests dan met scores op non-verbale intelligentietests. De correlaties met non-verbale intelligentietests variëren in het algemeen tussen .40 en .45.

Volgens Brus en Van Bergen (1973) bestaat er een nauwe samenhang tussen begrijpend lezen, algemene taalvaardigheid en denkvaardigheid of intelligentie. In het ijkingsonderzoek vonden zij in het vierde leerjaar een correlatie van .61 en .52 tussen de twee vormen van de Pinter-Durost Lagere School Test en een test voor begrijpend lezen nl. Schriftelijke Opdrachten. De samenhang tussen de betreffende intelligentietest en de test voor begrijpend lezen is middelmatig van aard en zou waarschijnlijk hoger zijn geweest wanneer een andere en betere intelligentietest was gebruikt.

Uit het voorgaande blijkt welke samenhang in het algemeen gevonden wordt tussen technisch lezen en begrijpend lezen, woordenschat en begrijpend lezen en intelligentie en begrijpend lezen. Opgemerkt moet worden dat in de vermelde onderzoeken gebruik is gemaakt van tests die begrijpend lezen in het algemeen meten en niet van tests die op specifieke taken van begrijpend lezen zijn gericht. Onderzoek waarbij de relatie wordt nagegaan tussen specifieke tests voor begrijpend lezen en tests voor technisch lezen, begrijpend lezen in het algemeen, woordenschat en intelligentie is schaars. Zoals in hoofdstuk I vermeld, ontwikkelden wij in een eerder onderzoek (Aarnoutse et al., 1977) enkele tests die betrekking hadden op het identificeren en afleiden van de hoofdgedachte van een tekst. In dit onderzoek werd onder meer de samenhang van twee hoofdgedachtetests met de Stilleestest van Bakker (1972) en de Woordenschattest van Stijnen (1975) vastgesteld. De correlaties tussen de testvormen waarbij de hoofdgedachte expliciet vermeld wordt en de Stilleestest bedroegen voor leerjaar 4, 5 en 6 respectievelijk .56, .56 en .53. De correlaties tussen dezelfde testvormen en de Woordenschattest waren .59, .59 en .58. De correlaties tussen de testvormen waarbij de hoofdgedachte moest worden afgeleid en de Stilleestest bedroegen voor leerjaar 4, 5 en 6 respectievelijk .47, .61 en .50. De correlaties tussen de laatst genoemde testvormen en de Woordenschattest waren voor leerjaar 4, 5 en 6 respectievelijk .52, .60 en .57. Vergelijkt men deze correlaties met de correlaties die Brus en Van Bergen (1973) in het vierde leerjaar vonden tussen de Stilleestest en de Schriftelijke Opdrachten ($r = .67$) en tussen de Stilleestest en de Woordenschattest ($r = .65$) dan zou men hieruit kunnen afleiden dat specifieke tests voor begrijpend lezen lager correleren met een test voor begrijpend lezen in het algemeen en voor woordenschat. Gezien het beperkt aantal proefpersonen ($n = \pm 200$ per leerjaar) en het feit dat de ontwikkelde hoofdgedachtetests voor verbetering vatbaar waren, is deze conclusie echter prematuur. Voorlopig gaan we ervan uit dat taakgerichte tests voor begrijpend lezen ongeveer dezelfde samenhang vertonen met maten voor begrijpend lezen en woordenschat als tests voor begrijpend lezen in het algemeen. Bovendien veronderstellen we dat taakgerichte tests voor begrijpend lezen ongeveer dezelfde samenhang vertonen met maten voor technisch lezen en intelligentie als tests voor begrijpend lezen in het algemeen.

Op basis van de voorgaande onderzoeksliteratuur en de veronderstelling dat

taakgerichte tests voor begrijpend lezen ongeveer dezelfde samenhang vertonen met maten voor technisch lezen, woordenschat en non-verbale intelligentie als tests voor begrijpend lezen in het algemeen verwachten we:^x

- a. een vrij lage correlatie tussen de drie taakgerichte tests en een maat voor technisch lezen;
- b. een aanzienlijke samenhang tussen de drie taakgerichte tests en een maat voor begrijpend lezen in het algemeen;
- c. een aanzienlijke samenhang tussen de drie taakgerichte tests en een maat voor leeswoordenschat;
- d. een middelmatige tot vrij lage samenhang tussen de drie taakgerichte tests en een maat voor non-verbale intelligentie.

Wat betreft de samenhang tussen de drie taakgerichte tests i.c. de Hoofdgedachtetest, de Bedoelingtest en de Redeneertest verwachten we dat deze tests in aanzienlijke mate zullen samenhangen. Deze verwachting is gebaseerd op de veronderstelling dat elk van deze tests in sterke mate een beroep doet op hogere cognitieve processen i.c. inferentieprocessen. Deze verwachting is niet in strijd met het eerder geformuleerde idee dat de drie leestaken i.c. tests verschillen wat betreft de gestructureerde en strategische kennis waarop bij de uitvoering van de taken i.c. de tests een beroep wordt gedaan. We verwachten niet dat de gemeenschappelijke variantie van de ware scores op deze tests van dien aard zijn dat deze tests in hoge mate hetzelfde meten.

Cloze tests

Tests waarbij de cloze procedure wordt toegepast, worden cloze tests genoemd. Zoals reeds eerder beschreven, bestaat de cloze procedure uit het aanvullen van een tekst waaruit woorden op een systematische wijze zijn weggelaten. Alle weggelaten woorden worden vervangen door open plekken van gelijke lengte en de proefpersonen wordt gevraagd de ontbrekende woorden in te vullen. Het meest gangbaar is elk n-de woord weg te laten, waarbij voor n meestal een

^x Een correlatie van .40 tot .60 wordt hier met de term middelmatig aangeduid. Een correlatie van .60 tot .70 wordt aanzienlijk genoemd. Een correlatie beneden .40 wordt gekarakteriseerd als laag of vrij laag, terwijl een correlatie boven .70 als hoog of vrij hoog wordt aangeduid.

getal tussen de vijf en de tien wordt gekozen. Het is ook mogelijk om woorden weg te laten door een tabel met random getallen te gebruiken. In beide gevallen worden woorden weggelaten zonder er op te letten tot welke woordsoort die woorden behoren. Het is echter ook mogelijk om slechts woorden van bepaalde woordsoorten weg te laten zoals functiewoorden of inhoudswoorden. In de standaard cloze procedure of free response procedure wordt van deze mogelijkheid meestal geen gebruik gemaakt. Een andere mogelijkheid is om woorden op een systematische wijze weg te laten en hierbij meerkeuze-items te construeren, waarbij de proefpersoon de keus krijgt uit een aantal alternatieven. Tests waarbij alternatieven worden aangeboden noemt men vaak multiple choice cloze tests.

In dit onderzoek worden twee typen cloze tests ontwikkeld nl. een free response cloze test waarbij de standaard procedure wordt toegepast en een multiple choice cloze test. In het vervolg zullen we eerst aandacht besteden aan onderzoek waarbij de free response procedure is toegepast en vervolgens aan onderzoek waarbij de multiple choice procedure is gebruikt.

De free response cloze procedure

De cloze procedure werd oorspronkelijk door Taylor (1953) ontwikkeld om de leesbaarheid van teksten te meten. Als door een groep personen meer van de weggelaten woorden in tekst A kan worden ingevuld dan in tekst B, dan is tekst A gemakkelijker dan tekst B voor deze groep. Toen Taylor de cloze procedure introduceerde, was het voor hem duidelijk dat het gemak waarmee de weggelaten woorden kunnen worden ingevuld niet alleen door de tekst wordt bepaald, maar ook door de bekwaamheid van de lezer. Cloze scores kunnen daarom niet alleen worden gebruikt om de leesbaarheid van verschillende teksten voor een bepaalde groep proefpersonen vast te stellen, maar ook om per tekst de verschillende proefpersonen te vergelijken wat betreft hun bekwaamheid de cloze taak uit te voeren. Taylor (1957) paste de cloze procedure op laatst genoemde wijze toe en noemde de totaalscore een maat voor tekstbegrip. Opgemerkt moet worden dat Taylor de cloze procedure voor het meten van begrijpend lezen eigenlijk opnieuw introduceerde. Uit een publicatie van 1929 blijkt bijvoorbeeld dat Van Veen in navolging van Ballard deze invulvorm al aanbeveelt voor het meten en oefenen van stil lezen.

Na de her-introductie van Taylor (1953, 1957) is vooral in de Verenigde

Staten veel onderzoek gedaan naar verschillende aspecten van de cloze procedure. Goede overzichtsartikelen zijn samengesteld door Rankin (1965), Potter (1968), Walter (1974) en in Nederland door Wesdorp (1978), terwijl Bickley et al. (1970) en Jongsma (1970) uitgebreide opsommingen geven van studies met betrekking tot de cloze procedure.

De meeste onderzoekers hebben tot nu toe bij het weglaten van woorden de standaard procedure toegepast, d.w.z. om de vijf tot tien woorden een woord weglaten. Taylor (1957) laat elke vijfde woord weg en wijst evenals Bormuth (1967) op de noodzaak van minstens 50 items per test.

Wat betreft de wijze van scoring van de cloze test is veel onderzoek gedaan. Walter (1974) komt na de bestudering van onderzoek tot de conclusie dat het best gekeken kan worden of het weggelaten woord is ingevuld. Het geheel of gedeeltelijk goed rekenen van synoniemen en dergelijke verhoogt wel de gemiddelde totaalscore, maar beïnvloedt de validiteits- en betrouwbaarheidscoefficienten niet of nauwelijks. Wesdorp (1978) pleit eveneens voor een exacte scoringswijze. Wijnstra (1977) beschrijft een onderzoek van Levin waarin onder meer twee scoringswijzen met elkaar worden vergeleken nl. de exacte scoringswijze en de semantische scoringswijze. De correlatie tussen de totaalscores verkregen na beide wijzen van scoring is hoog nl. .95. De validiteits- en betrouwbaarheidscoefficienten vielen bij de semantische scoring iets hoger uit. Wijnstra (1977) past dan ook een semantische scoringswijze toe, waarbij twee onafhankelijke beoordelaars een overeenstemmingspercentage bereiken van 96%.

Uit verschillende onderzoeken blijkt dat de correlaties hoog zijn wanneer voor zowel de free response cloze test als de multiple choice cloze test dezelfde teksten worden gebruikt. Wijnstra (1977) vond een correlatie van .86 tussen beide typen cloze tests.

Er is ook onderzoek gedaan naar het systematisch weglaten van een bepaald percentage woorden van één of meer woordsoorten. Volgens Hauwermeiren (1975) zijn cloze tests met structurele deleties (functiewoorden) gemakkelijker dan cloze tests met semantische deleties (inhoudswoorden). Rankin (1972) maakt onderscheid tussen lexicaal begrip waarbij inhoudswoorden worden weggelaten en structureel begrip, waarbij functiewoorden worden weggelaten. Volgens O'Reilly et al. (1976) meten cloze tests waarbij inhoudswoorden zijn weggelaten een beperkter aspect van begrijpend lezen dan wanneer elk n-de woord

wordt weggelaten.

Er is veel onderzoek gedaan om de validiteit van de cloze score als maat voor begrijpend lezen te bepalen. Over het algemeen wordt deze maat gecorrigeerd met gegevens verkregen met behulp van gestandaardiseerde tests voor begrijpend lezen. Rankin (1965) concludeert uit zijn overzicht van onderzoek naar het gebruik van de cloze procedure voor het meten van begrijpend lezen, dat de correlaties tussen scores op gestandaardiseerde leestests en scores op cloze tests waarin elk n-de woord is weggelaten, over het algemeen hoog zijn. Tot dezelfde conclusie komen Bormuth (1967) en Jongsma (1970). Volgens Bickley et al. (1970) is de cloze test een geschikte test voor het meten van begrijpend lezen in het algemeen. Weaver en Kingston (1963) komen op grond van hun onderzoeksgegevens echter tot een geheel andere conclusie nl. dat cloze tests niet geschikt zijn om begrijpend lezen te meten. Volgens Bormuth (1969) zijn de afwijkende resultaten van Weaver en Kingston onder andere te verklaren door de selecte steekproef nl. studenten en de gestandaardiseerde tests die werden gebruikt.

In verband met de vaststelling van de validiteit van de te ontwikkelen free response cloze test zijn drie onderzoeken van belang nl. het onderzoek van Bormuth (1969), Wijnstra (1977) en Aarnoutse et al. (1977).

Bormuth (1969) voerde een onderzoek uit naar de factor validiteit van standaard cloze tests in leerjaar vier, vijf en zes van het lager onderwijs. Hij construeerde negen cloze tests met teksten van ongeveer 250 woorden, waaruit elk vijfde woord werd weggelaten. Met dezelfde negen teksten ontwikkelde hij zeven subtests voor begrijpend lezen. Deze subtests waren bedoeld om verschillende aspecten van begrijpend lezen te meten nl. woordenschat, het begrijpen van expliciet vermelde feiten, het begrijpen van volgorden van gebeurtenissen, het begrijpen van causale relaties, het begrijpen van hoofdgedachten en van de bedoeling van een schrijver. De items van de subtests werden door drie deskundigen beoordeeld en in enkele klassen beproefd. Daarna werden eerst de negen cloze tests bij 150 leerlingen afgenomen en vervolgens bij dezelfde leerlingen de zeven subtests. De split-half betrouwbaarheidscoëfficiënten van de cloze tests lagen tussen .84 en .88 (bij 50 items per test); de coëfficiënten van zes subtests lagen tussen .71 en .93. De test voor het meten van de hoofdgedachte had een lage betrouwbaarheid nl. .43. De eerste principale component van de correlaties tussen de cloze tests en de

andere tests verklaarde 77% van de variantie. De ladingen op de eerste factor benaderen alle waarden van de betrouwbaarheidsindex i.c. de wortel uit de split-half betrouwbaarheidscoëfficiënt. Op basis van deze gegevens concludeert Bormuth dat cloze tests praktisch hetzelfde meten als de conventionele meerkeuze-tests voor begrijpend lezen.

Wijnstra (1977) construeerde een nederlandstalige cloze test en nam deze af bij 102 leerlingen van leerjaar 4, 5 en 6 van een school. Hij vond een betrouwbaarheidscoëfficiënt van .89. De test-hertest betrouwbaarheidscoëfficiënt bedroeg .93, terwijl de correlatie met een andere cloze test .70 bleek te zijn. In het kader van het project Friesland werden onder andere bij 120 Utrechtse leerlingen van het zesde leerjaar verschillende tests afgenomen waaronder de bovengenoemde cloze test, de Woordenschattest van Stijnen (1975), de Standard Progressive Matrices van Raven (1960) en het onderdeel Stillezen van de Eindtoets Basisonderwijs van het CITO. De correlaties van de cloze test van Wijnstra met de bovengenoemde woordenschattest, de intelligentietest en het onderdeel Stillezen bedroegen respectievelijk .72, .42 en .69. Uit deze en andere gegevens leidt Wijnstra af dat de cloze test een zeer bruikbare maat is voor schriftelijke taalvaardigheid.

Zoals reeds in hoofdstuk I vermeld, ontwikkelden wij in het onderzoek van 1976-1977 twee cloze tests (Aarnoutse et al., 1977). Eén cloze test was bestemd voor het derde en vierde leerjaar en één test voor het vijfde en zesde leerjaar (elke test bevatte 50 items). Het aantal leerlingen waarbij deze tests werd afgenomen, bedroeg ruim 200 per leerjaar. De KR-20 betrouwbaarheid van de eerste cloze test, die uit twee verschillende vormen bestond, varieerde van .72 tot .84. De KR-20 betrouwbaarheid van de tweede cloze test voor leerjaar 5 en 6 varieerde van .68 tot .70. De correlaties tussen deze cloze tests en de Stilleestest van Bakker (1972) bedroegen in de leerjaren drie tot en met zes respectievelijk .58, .52, .58 en .59. De correlaties tussen deze cloze tests en de Woordenschattest van Stijnen (1975) bedroegen voor dezelfde leerjaren respectievelijk .67, .61, .60 en .59. Uit deze gegevens blijkt dat in dit onderzoek niet zo hoge betrouwbaarheidscoëfficiënten werden gevonden als in het onderzoek van Wijnstra en dat ook de correlaties met de Woordenschattest lager zijn.

Het feit dat in onderzoek vaak vrij hoge correlaties worden gevonden tussen free response cloze tests en meerkeuze-tests voor begrijpend lezen is

uiteraard geen voldoende bewijs dat cloze tests inderdaad begrijpend lezen meten. Volgens Jongsma (1970) is er weinig bekend omtrent de processen die zich afspelen bij het uitvoeren van een cloze test. Volgens Carroll (1972) lijken cloze scores voor een groot deel afhankelijk te zijn van wat hij noemt de lokale redundantie van een tekst, d.w.z. de mate waarin de onmiddellijke omgeving - in het algemeen dezelfde zin - van een ontbrekend woord aanwijzingen bevat om dat woord te vinden. Cloze scores zijn volgens hem waarschijnlijk meer afhankelijk van het opsporen van grammaticale dan van semantische aanwijzingen. Rankin (1974) noemt in dit verband de standaard cloze procedure de structurele cloze.

De multiple choice cloze procedure

Een belangrijk voordeel van de free response cloze test is de eenvoudige constructieprocedure. De scoring vraagt echter veel tijd, tenminste als een semantische scoring wordt toegepast. Om deze en andere - meer fundamentele - redenen zijn sommige onderzoekers er toe overgegaan om de standaard cloze procedure zodanig aan te passen dat in plaats van open plekken in een tekst een aantal alternatieven wordt aangeboden, waarvan er telkens één gekozen moet worden.

In het volgende zullen we aandacht besteden aan de multiple choice cloze tests zoals die door O'Reilly, Schuder en Kidder (1976) zijn ontwikkeld en aan het validiteitsonderzoek dat in verband met deze tests is uitgevoerd. De reden dat we hier de constructie en validering van de tests van O'Reilly c.s. bespreken, is vooral gelegen in het feit dat de multiple choice cloze test in het onderhavige onderzoek voor een deel is geconstrueerd volgens de procedure van O'Reilly c.s.

O'Reilly, Schuder en Kidder (1976) hebben multiple choice cloze tests ontwikkeld met de bedoeling het letterlijk begrijpen van lezen te meten. Volgens hen meten de gestandaardiseerde tests voor begrijpend lezen vooral de hogere cognitieve processen zoals inferentiële processen. Tests die "minimum comprehension" bij lezen meten, zijn er volgens Kidder en Schuder (1975) niet.

Kidder en Schuder (1975) hebben geprobeerd in hun literatuurstudie een bruikbare omschrijving te vinden van het begrip "literal comprehension". Uit hun verslag blijkt dat een dergelijke omschrijving niet voorhanden is en dat de definitie van Katz van de letterlijke betekenis van een zin nog het best

bruikbaar is. Ook het onderscheid van Carroll (1972) is volgens hen verhelderend. Zoals we gezien hebben, onderscheidt Carroll twee fasen in het proces van het begrijpen van gesproken en geschreven taal nl. het begrijpen van de linguïstische informatie en het relateren van deze informatie aan een bredere context. Kidder en Schuder (1975) omschrijven letterlijk begrijpen als "the act or process of grasping the explicit, societally standardized meaning of a verbal message" (pag. 19). Volgens O'Reilly en Streeter (1977) heeft het letterlijk begrijpen betrekking op het begrijpen van de linguïstische informatie of "language comprehension per se" (pag. 46). De bovenstaande omschrijvingen zijn weinig exact en theoretisch nauwelijks gefundeerd. Het is overigens ook de vraag of het betreffende begrip theoretisch gefundeerd kan worden. In hoofdstuk II hebben we gezien dat inferentiële processen een essentiële rol spelen bij het begrijpen van tekstuele informatie.

Op grond van hetgeen verstaan wordt onder letterlijk begrijpen en rekening houdend met de voor- en nadelen van de standaard cloze procedure ontwikkelden O'Reilly c.s. een aangepaste vorm van de cloze procedure met de volgende kenmerken: (a) alleen inhoudswoorden worden weggelaten, (b) de hoeveelheid weg te laten inhoudswoorden bedraagt 20% van het totaal, (c) de afleiders zijn grammaticaal juist, maar passen semantisch niet in de tekst. Wat betreft de inhoudswoorden (zelfstandige naamwoorden, werkwoorden, bijvoeglijke naamwoorden en bijwoorden) wordt aangenomen dat deze de belangrijkste dragers zijn van de informatie in de tekst. Het percentage weg te laten inhoudswoorden komt overeen met het percentage woorden dat weggelaten wordt als elk vijfde woord wordt geschrapt. Aanvullende regels worden opgesteld om tot een objectieve weglatingsprocedure te komen. Doordat de afleiders grammaticaal juist zijn, kan niet op grond van de grammaticale vorm het juiste antwoord worden gekozen; de semantische inhoud is hiervoor bepalend.

O'Reilly, Schuder en Kidder (1976) beschrijven uitvoerig op welke wijze zij multiple choice cloze tests hebben ontwikkeld voor de leerjaren 1 tot en met 9. Teksten van 70-80 woorden werden geselecteerd uit de leerstof voor de verschillende leerjaren en voor verschillende leergebieden. Van elke tekst werd met behulp van leesbaarheidsformules een leesbaarheidsindex bepaald. Vervolgens werden woorden op een systematische wijze weggelaten. Voor de keuze van de afleiders werd gebruik gemaakt van woordenlijsten en woordfrequentie-

tielijsten. De afleiders werden eveneens op een systematische wijze gegene-reerd.

Uit de publicaties van O'Reilly c.s. blijkt dat op de bovenbeschreven wijze betrouwbare testvormen zijn te construeren. Zo was de mediaan van de KR-20 betrouwbaarheid van 12 testvormen van elk 60 items in de leerjaren vier tot en met zes tesamen .96 (range .94 tot .97).

Voor het onderzoek naar de validiteit van de multiple choice cloze tests (MCCT's) werden de scores op deze tests vergeleken met een andere maat voor letterlijk begrijpen, een maat voor andere aspecten van begrijpend lezen, en een maat voor intelligentie en testvaardigheid. Als andere maat voor letterlijk begrijpen werden verschillende testvormen ontwikkeld volgens de "wh-item" testprocedure van Bormuth (1970). Als variabelen die sterk met het construct letterlijk begrijpen samenhangen, werden enkele subtests uit de California Achievement Test-batterij gekozen (CAT). Als variabelen waarmee letterlijk begrijpen veel minder sterk samenhangt (divergente validiteit), werden de overige subtests van de CAT-batterij gekozen evenals een maat voor verbale en non-verbale intelligentie, gemeten met de Short Form Test of Academic Aptitude.

Op de correlaties tussen de verschillende variabelen werden door O'Reilly en Streeter (1977) voor het eerste leerjaar, het tweede en derde leerjaar samen en het vierde tot en met het zesde leerjaar diverse principale componentenanalyses uitgevoerd met verschillende rotaties. Volgens de onderzoekers lijken de resultaten de conclusie te ondersteunen dat de MCCT voor een deel een maat is voor een beperkt aspect van begrijpend lezen dat grotendeels onafhankelijk is van intelligentie. Dit aspect lijkt volgens hen te interpreteren als het begrijpen van de strikt letterlijke betekenis binnen zinnen en zinsdelen. Uit de analyses blijkt echter dat de scores op de MCCT nog een tweede en zelfs een derde component bevatten naast het strikt letterlijk begrijpen. De conclusie als zou de MCCT een aspect van begrijpend lezen meten dat grotendeels onafhankelijk is van intelligentie, is bovendien onjuist. De correlatie tussen de MCCT-scores en verbale en non-verbale intelligentie blijkt in de leerjaren 4, 5 en 6 aanzienlijk tot middelmatig te zijn nl. .62 en .52. Uit deze correlaties kan worden afgeleid dat de MCCT geen zuivere maat is voor wat letterlijk begrijpen wordt genoemd. Naar onze mening is de MCCT eerder een maat voor begrijpend lezen in het algemeen.

Op basis van de onderzoeksliteratuur met betrekking tot de free response cloze procedure kunnen we een aantal verwachtingen formuleren over mogelijke verbanden tussen de Free Response Cloze Test (FRCT) en maten voor lezen, woordenschat en non-verbale intelligentie. We verwachten dat de FRCT aanzienlijk zal correleren met de taakgerichte tests i.c. de Hoofdgedachtetest, de Bedoelingtest en de Redeneertest. Deze verwachting is gebaseerd op de veronderstelling dat bij de FRCT inferentiële processen een belangrijke rol spelen evenals bij de taakgerichte tests. Verder verwachten we dat de FRCT aanzienlijk zal correleren met maten voor begrijpend lezen in het algemeen en woordenschat. Een middelmatige tot lage samenhang verwachten we tussen de FRCT en een maat voor technisch lezen en voor non-verbale intelligentie.

Wat betreft de multiple choice cloze test (MCCT) verwachten we ongeveer dezelfde correlaties met maten voor lezen, woordenschat en non-verbale intelligentie als die welke voor de FRCT worden gevonden. Als we uitgaan van de stelling van O'Reilly c.s. dat de MCCT voor een groot deel een maat is voor letterlijk begrijpen, dan moet de MCCT middelmatig tot laag correleren met de taakgerichte tests en met de FRCT.

In dit hoofdstuk wordt na een inleiding het doel van het onderzoek nader geconcretiseerd in de vorm van vraagstellingen. Daarna volgt de opzet van het onderzoek en wordt een beschrijving gegeven van de ontwikkeling van de tests en de programma's. Vervolgens komt de selectie van de scholen aan de orde evenals een vergelijking van de experimentele en controlescholen wat betreft het leesonderwijs. Tenslotte wordt een beschrijving gegeven van de dataverzameling en van de wijze waarop het experiment is voorbereid en uitgevoerd.

4.1. Inleiding

Voor een goed begrip van de vraagstellingen en de opzet van het onderzoek geven we hier in het kort een beschrijving van het doel van het onderzoek en van enkele activiteiten die tot dusverre zijn ondernomen.

Zoals uit hoofdstuk I blijkt, bestaat het onderzoek uit twee onderscheiden delen nl. een descriptief gedeelte en een experimenteel gedeelte. Het descriptief gedeelte van het onderzoek heeft vooral tot doel de betrouwbaarheid en validiteit vast te stellen van vijf tests voor het begrijpend lezen van teksten nl. drie taakgerichte tests en twee cloze tests. Verder heeft het descriptief gedeelte tot doel de prestaties en de verandering in prestaties vast te stellen van leerlingen van het vierde leerjaar op de betreffende tests voor begrijpend lezen. Het experimenteel gedeelte van het onderzoek heeft tot doel het effect vast te stellen van drie programma's (leereenheden) die elk afzonderlijk zijn gericht op een betere uitvoering van één bepaalde taak voor het begrijpend lezen van teksten. Bovendien wordt in het experimentele gedeelte de realiteitswaarde vastgesteld van de onderscheiding van drie taken op het gebied van begrijpend lezen.

In hoofdstuk II hebben we een beschrijving gegeven van de selectie en specificatie van de drie taken die in dit onderzoek centraal staan en waarvoor taakgerichte tests en -programma's worden ontwikkeld. Deze taken zijn: het afleiden van de hoofdgedachte uit een informatieve tekst (de taak Hoofdgedachte), het afleiden van de bedoeling van de hoofdpersoon uit een fictieve, verhalende tekst (de taak Bedoeling) en het trekken van conclusies uit

logische redeneringen (de taak Redeneren). De drie taakgerichte tests worden aangeduid als Hoofdgedachtetest, Bedoelingstest en Redeneertest. De taakgerichte programma's worden aangeduid als het programma Hoofdgedachte, Bedoeling en Redeneren.

In hoofdstuk III wordt onder meer een verantwoording gegeven van de opzet van de drie taakgerichte programma's. Bovendien wordt vermeld dat in navolging van Chapman (1971) de realiteitswaarde van de onderscheiding van de drie geselecteerde taken via een experimentele opzet zal worden vastgesteld en dat in navolging van Davis (1968) bij de ontwikkeling van de taakgerichte tests elk item gebaseerd zal worden op één afzonderlijke tekst. Verder wordt in dit hoofdstuk beschreven wat verstaan moet worden onder taakgerichte tests en cloze tests met name de te ontwikkelen Free Response Cloze Test (FRCT) en de Multiple Choice Cloze Test (MCCT). Tenslotte worden verwachtingen geformuleerd omtrent mogelijke verbanden tussen de vijf te ontwikkelen tests en maten voor technisch lezen, begrijpend lezen, woordenschat en non-verbale intelligentie.

4.2. Vraagstellingen

De vraagstellingen van het descriptieve gedeelte van het onderzoek kunnen als volgt worden geformuleerd:

- a. Welke prestaties leveren leerlingen van het vierde leerjaar op drie taakgerichte tests nl. de Hoofdgedachtetest, de Bedoelingstest en de Redeneertest? Welke prestaties leveren dezelfde leerlingen op twee cloze tests die betrekking hebben op begrijpend lezen in het algemeen nl. de Free Response Cloze Test en de Multiple Choice Cloze Test?
- b. Hoe betrouwbaar zijn de vijf onder a. genoemde tests en de parallelvormen van deze tests?
- c. Welke samenhang bestaat er tussen de vijf tests genoemd onder a. en wat is de samenhang tussen deze tests en gestandaardiseerde tests voor technisch lezen, begrijpend lezen, woordenschat en non-verbale intelligentie? Het antwoord op deze vraag is vooral van belang in verband met de validiteit van de vijf tests.
- d. Welke verandering in prestaties treedt op bij de vijf onder a. genoemde tests over drie meetmomenten?

De bovengenoemde vraagstellingen worden in hoofdstuk V nader gepreciseerd. Wat betreft de vraagstelling genoemd onder c. zullen verwachtingen worden geformuleerd.

De vraagstellingen van het experimenteel gedeelte van het onderzoek kunnen als volgt worden geformuleerd:

- a. Wat is op korte en lange termijn het effect van elk van de drie taakgerichte programma's op de beheersing van de desbetreffende taken, gemeten met de daarmee corresponderende taakgerichte tests?
- b. Wat is op korte en lange termijn het effect van elk programma op de beheersing van die taken waarop de beide andere programma's zijn gericht? Wat is bijvoorbeeld het effect van het programma Hoofdgedachte op de beheersing van de redeneertaak en de bedoelingtaak? Deze vraag vormt een nadere concretisering van de vraag naar de realiteitswaarde van de onderscheiding van de drie geselecteerde taken op het gebied van begrijpend lezen.
- c. Wat is op korte en lange termijn het effect van elk programma op de ontwikkeling van het begrijpend lezen in het algemeen? Het begrijpend lezen in het algemeen wordt gemeten met behulp van de Free Response Cloze Test en de Multiple Choice Cloze Test.

De vraagstellingen van het experimenteel gedeelte worden in hoofdstuk V nader omschreven in de vorm van hypothesen.

4.3. Opzet van het onderzoek

Alvorens de opzet van het onderzoek te beschrijven, vermelden we eerst welke tests in verband met het onderzoek naar de validiteit van de te ontwikkelen tests werden geselecteerd.

Zoals eerder vermeld, wordt in dit onderzoek de samenhang vastgesteld tussen de te ontwikkelen tests en gestandaardiseerde tests voor technisch lezen, begrijpend lezen, woordenschat en non-verbale intelligentie. Als maat voor technisch lezen werd de Eén-Minuut-Test van Brus en Voeten (1973) gekozen. Deze test pretendeert de vaardigheid in het ontsleutelen van gedrukte woorden te meten. De test bestaat uit een kaart met 116 niet-samenhangende woorden. De ruwe score wordt gevormd door het aantal goed gelezen woorden in één minuut. Als maat voor begrijpend lezen werden twee tests geselecteerd nl. de

Schriftelijke Opdrachten van Brus en Van Bergen (1973) en de Stilleestest van Bakker (1972). De test Schriftelijke Opdrachten heeft betrekking op het begrijpen van de betekenis van korte aanwijzingen (opdrachten) gevat in één of twee zinnen. De test bestaat uit 32 in moeilijkheid opklimmende gedrukte opdrachten, die moeten worden uitgevoerd. De Stilleestest heeft betrekking op het begrijpen van de betekenis van langere teksten. De test bestaat uit 12 verhaaltjes, waarover een of meer vragen worden gesteld waarop het juiste antwoord uit vijf alternatieven gekozen moet worden. De versie voor het vierde leerjaar bevat in totaal 30 items. Als maat voor woordenschat werd de Woordenschattest van Stijnen (1975) gekozen. Deze test beoogt de lees-woordenschat (receptief-schriftelijke woordenschat) van kinderen van het derde tot en met zesde leerjaar te meten. De test voor het vierde leerjaar bestaat uit 61 items. Bij elk item moeten de kinderen uit vier gegeven omschrijvingen van een woord die omschrijving kiezen, die het best overeenkomt met de betekenis van een woord. Omdat er relatief weinig onderzoek in Nederland is verricht naar de relatie tussen begrijpend lezen en non-verbale intelligentie (in tegenstelling tot verbale intelligentie) wordt in dit onderzoek de samenhang vastgesteld tussen tests voor begrijpend lezen en non-verbale intelligentie. Als maat voor non-verbale intelligentie werd de Standard Progressive Matrices van Raven (1958) gekozen. Deze test, die door Raven als een "test of observation and clear thinking" (1958, pag. 2) wordt gekarakteriseerd, bestaat uit 5 delen van elk 12 items die opklimmen in moeilijkheid. Elk item bestaat uit een patroon waarvan een deel ontbreekt. De proefpersonen moeten uit een aantal alternatieven (variërend van 6 tot 8) het juiste deel selecteren.

Vanaf het begin van het onderzoek was duidelijk dat vanwege praktische en financiële redenen het niet mogelijk was het onderzoek uit te voeren op scholen die een aselechte steekproef vormen uit de Nederlandse lagere scholen. Besloten werd om in totaal acht grote lagere scholen op een systematische wijze te selecteren uit een vooraf bepaald gebied en random twee groepen van scholen te vormen. Beide groepen van scholen bestaan uit vier scholen met minimaal twee vierde klassen per school. Op de ene groep scholen worden de experimentele programma's gegeven, terwijl op de andere groep scholen geen experimentele programma's worden gegeven. De eerste groep scholen wordt gemakshalve aangeduid met de term experimentele scholen en de tweede groep met de

term controlescholen. In par. 4.6. van dit hoofdstuk geven we een gedetailleerde beschrijving van de selectie en verdeling van de scholen.

Het onderzoek begint met een voormeting. Hieraan nemen alle scholen, zowel de experimentele als de controlescholen, deel. Bij de voormeting worden de volgende tests afgenomen:

- a. de Schriftelijke Opdrachten van Brus en Van Bergen (vierde leerjaar);
- b. de Woordenschattest van Stijnen (vierde leerjaar);
- c. de Stilleestest van Bakker (vierde leerjaar);
- d. de Eén-Minuut-Test van Brus en Voeten;
- e. de Standard Progressive Matrices van Raven;
- f. vorm A van de Multiple Choice Cloze Test (MCCT);
- g. vorm A van de Free Response Cloze Test (FRCT);
- h. vorm A van de Hoofdgedachtetest (HGD);
- i. vorm A van de Redeneertest (RED);
- j. vorm A van de Bedoelingtest (BED).

De tests genoemd onder a tot en met e zijn gestandaardiseerde tests, waarvan de eerste vier voor Nederland zijn geijkt. De tests genoemd onder f tot en met j zijn tests voor begrijpend lezen die, zoals reeds eerder vermeld, in het kader van dit onderzoek worden ontwikkeld. In par. 4.4. van dit hoofdstuk geven we een beschrijving van de ontwikkeling van deze tests.

Tijdens de voormeting worden enkele gegevens van de leerlingen verzameld zoals leeftijd, geslacht, sociaal milieu en rapportcijfers voor lezen, taal en rekenen.

Na de voormeting worden de leerlingen van elke experimentele school in vier groepen verdeeld, die wat betreft leesbekwaamheid gelijk zijn samengesteld. Aan elke groep wordt één van de vier programma's (Hoofdgedachte, Bedoeling, Redeneren en het Controleprogramma) onderwezen. Elk programma bestaat uit 10 lessen en wordt gedurende vijf weken met een frequentie van twee lessen per week gegeven door student-assistenten (instructeurs). In par. 4.5. van dit hoofdstuk geven we een beschrijving van de ontwikkeling van de experimentele programma's. In par. 4.8. volgt een beschrijving van het controleprogramma.

Na afloop van de programma's volgt de nameting, die zowel op de experimentele scholen als op de controlescholen gehouden wordt. De nameting bestaat uit de B-vormen van de tests vermeld onder f tot en met j.

Het onderzoek wordt enige tijd later afgesloten met een retentiemeting. Ook deze meting wordt op zowel de experimentele als de controlescholen uitgevoerd. De retentiemeting bestaat uit dezelfde tests als de nameting met dien verstande dat de A-vormen van de voormeting weer gebruikt worden.

De opzet van het onderzoek is schematisch weergegeven in figuur 4.1.

Figuur 4.1. Opzet van het onderzoek.

school	Experimentele scholen								Controlescholen							
	a		b		c		d		e		f		g		h	
4e klas	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
nov/ dec	Voormeting								Voormeting							
jan/ febr	H	B	R	C	H	B	R	C	H	B	R	C	H	B	R	C
				P				r								a
maart	Nameting								Nameting							
mei/ juni	Retentiemeting								Retentiemeting							

H = programma Hoofdgedachte
 B = programma Bedoeling
 R = programma Redeneren
 C = Controleprogramma

Na deze globale beschrijving van de opzet van het onderzoek wordt eerst ingegaan op de opzet binnen de vier experimentele scholen en vervolgens op de opzet voor zover die betrekking heeft op de vergelijking van de experi-

mentele scholen en de controlescholen.

Schematisch voorgesteld ziet de opzet van het experiment binnen de vier experimentele scholen er als volgt uit:

	RO_1	X_1	O_2	O_3
Experimentele groep:	RO_1	X_2	O_2	O_3
	RO_1	X_3	O_2	O_3
Controlegroep:	RO_1	X_4	O_2	O_3

waarin:

R = randomisatie binnen elke experimentele school

O_1 = voormeting: de tests genoemd onder a tot en met j (pag. 119)

X_1 = programma Hoofdgedachte

X_2 = programma Bedoeling

X_3 = programma Redeneren

X_4 = Controleprogramma

O_2 = nameting: de B-vormen van de tests genoemd onder f tot en met j
(pag. 119)

O_3 = retentiemeting: de tests genoemd onder f tot en met j (pag. 119)

In elke experimentele school worden groepjes van telkens vier leerlingen gevormd, gematched op leesbekwaamheid. Binnen de door de matchingsprocedure gestelde grenzen worden de leerlingen van elk viertal random over vier groepen verdeeld. Deze vier groepen zijn wat betreft leesbekwaamheid gelijk samengesteld. De experimentele groep bestaat uit drie groepen leerlingen die geconfronteerd worden met één van de te ontwikkelen programma's nl. het programma Hoofdgedachte, Bedoeling of Redeneren. De controlegroep bestaat uit één van de vier bovengenoemde groepen; deze groep wordt geconfronteerd met een leesprogramma dat geen betrekking heeft op de te ontwikkelen programma's. Dit programma wordt in het vervolg aangeduid als controleprogramma. Met name om de invloed van variabelen als school en speciale aandacht (Hawthorne-effect) op de testprestaties te controleren, is deze controlegroep

geformeerd. Door veranderingen in de experimentele groep te vergelijken met die in de controlegroep kan het effect van de te ontwikkelen programma's (Hoofdgedachte, Redeneren, Bedoeling) beter worden geïsoleerd. Voor nadere gegevens met betrekking tot de samenstelling van de experimentele groep en de controlegroep verwijzen we naar par. 4.8.

Daar voor de uitvoering van de programma's speciale instructeurs worden ingeschakeld, heeft de variabele "klasseleerkracht" geen invloed op de testprestaties van de leerlingen. Om te voorkomen dat er een contaminatie optreedt tussen instructeur en programma of tussen instructeur en school worden de vier programma's door vier verschillende instructeurs op vier verschillende scholen gegeven. Voor nadere gegevens over de verdeling van de instructeurs over de scholen en de programma's verwijzen we eveneens naar par. 4.8.

De opzet van het experiment binnen de experimentele scholen kan worden gekarakteriseerd als een "true experimental pretest-posttest control group design" (Campbell & Stanley, 1971).

Een vergelijking van de condities van de experimentele scholen en de controlescholen ziet er schematisch als volgt uit:

Experimentele scholen:	O_1	X	O_2	O_3
Controlescholen:	O_1		O_2	O_3

waarin:

O_1 = voormeting: de tests genoemd onder a tot en met j (pag. 119)

X = het programma Hoofdgedachte, Bedoeling, Redeneren of het
Controleprogramma

O_2 = nameting: de B-vormen van de tests genoemd onder f tot en met
j (pag. 119)

O_3 = retentiemeting: de tests genoemd onder f tot en met j (pag. 119)

Zoals reeds eerder vermeld, krijgen de leerlingen van de experimentele scholen gedurende vijf weken onderwijs in één van de vier programma's (drie te ontwikkelen programma's en een controleprogramma). De leerlingen van de con-

trolescholen krijgen geen van de vier programma's aangeboden; zij volgen het reguliere leesonderwijs.

Deze experimentele opzet maakt het mogelijk het effect van de te ontwikkelen programma's en van het controleprogramma te vergelijken met het effect van het reguliere leesonderwijs. Een dergelijke opzet kan worden gekarakteriseerd als een quasi-experimenteel "non-equivalent control group design" (Campbell & Stanley, 1971). Hoewel de scholen uit één gemeenschappelijke populatie aselekt verdeeld worden in een groep experimentele scholen en een groep controlescholen, kan gezien de beperkte steekproef niet verwacht worden dat de beide groepen van scholen op alle relevante variabelen gelijk zijn. Nagegaan zal worden in hoeverre de scores op de voormeting voor beide groepen van scholen overeenkomen. Bovendien zal worden nagegaan in hoeverre de groepen scholen vergelijkbaar zijn wat betreft het leesonderwijs. Voor nadere gegevens omtrent de twee laatst genoemde punten wordt verwezen naar par. 5.2. van hoofdstuk V en naar par. 4.6. van dit hoofdstuk.

4.4. Ontwikkeling van de tests

In deze paragraaf wordt in het kort een beschrijving gegeven van de vijf tests die in het kader van dit onderzoek zijn ontwikkeld evenals de parallelvormen daarvan. Voor meer informatie verwijzen we naar Deelrapport 4 (Aarnoutse et al., 1979).

4.4.1. De Hoofdgedachtetest

Taak en kenmerken van de teksten

In hoofdstuk II zijn de belangrijkste resultaten vermeld van de literatuurstudie, die in verband met de specificatie van de hoofdgedachtetaak is uitgevoerd. Bovendien is een poging gedaan de taak zo goed mogelijk te omschrijven. De hoofdgedachte wordt omschreven als die uitspraak die in meer algemene termen weergeeft wat de specifieke uitspraken tezamen tot uitdrukking brengen. Onder specifieke uitspraken worden die mededelingen verstaan die de functie hebben van illustrerende bijzonderheden, voorbeelden of uitwerkingen van de meer algemene uitspraak i.c. de hoofdgedachte. Het afleiden van de hoofdgedachte wordt omschreven als het vinden van een uitspraak over

het thema van de tekst zó dat verschillende mededelingen fungeren als illustrerende bijzonderheden, voorbeelden of uitwerkingen hiervan. Met de term "afleiden" wordt niet bedoeld dat de leerlingen de hoofdgedachte zelf genereren of formuleren. Na de tekst en de vraag "Welke zin geeft het beste aan waar het hele verhaal over gaat?" moeten de leerlingen het juiste antwoord i.c. de hoofdgedachte uit een viertal alternatieven selecteren.

Na de omschrijving van de taak werden criteria geformuleerd waaraan de teksten moesten voldoen. Deze criteria waren:

- a. de teksten moeten informatief van aard zijn. Dit wil zeggen dat ze beschrijvend moeten zijn en niet fictief, overredend of directief;
 - b. de teksten moeten over één bepaald onderwerp handelen;
 - c. de teksten moeten een aantal uitspraken over het onderwerp bevatten, die tezamen één bepaalde gedachte tot uitdrukking brengen. Deze uitspraken dienen de functie te hebben van illustrerende bijzonderheden, voorbeelden of uitwerkingen van één bepaalde niet expliciet geformuleerde gedachte;
 - d. de teksten moeten impliciet een hoofdgedachte bevatten en niet te veel aanwijzingen in de richting van de hoofdgedachte;
 - e. de teksten moeten niet te veel moeilijke woorden bevatten en geen ingewikkelde zinsconstructies; moeilijke woorden moeten zo veel mogelijk in de tekst zelf worden verklaard;
 - f. de teksten moeten een omvang hebben van ± 90 tot ± 150 woorden;
 - g. de teksten moeten geselecteerd of bewerkt worden uit het leesmateriaal van kinderen van leerjaar 4 en 5;
 - h. de teksten moeten over een onderwerp handelen dat de kinderen van het vierde leerjaar enigszins bekend is;
 - i. de teksten moeten betrekking hebben op verschillende leer- of vakgebieden.
- Ter verduidelijking van het bovenstaande volgt nu een item uit de definitieve versie van de Hoofdgedachtetest (Vorm B, item 18).

Door een heleboel kleinere en grotere buisjes stroomt het bloed naar alle delen van ons lichaam. Een sterke pomp, ons hart, zorgt ervoor dat het bloed steeds blijft stromen. Bloed vervoert voedingsstoffen en zuurstof naar de kleinste delen van ons lichaam. Bovendien haalt het de afvalstoffen op. Het beschermt ons ook tegen allerlei ziekten, die ons dagelijks bedreigen. Bloed bevat verder stoffen die ervoor zorgen dat wonden weer snel genezen.

18. Welke zin geeft het beste aan waar het hele verhaal over gaat?

1. Het hart zorgt ervoor dat het bloed blijft stromen.
2. Bloed vervult verschillende taken in ons lichaam.
3. Bloed vervoert allerlei stoffen door ons lichaam.
4. Bloed stroomt naar alle delen van ons lichaam.

In deze tekst worden meerdere uitspraken gedaan over bloed. Zo wordt verteld dat bloed (a) door het lichaam stroomt, (b) voedingsstoffen en zuurstof vervoert, (c) afvalstoffen ophaalt, (d) het lichaam beschermt tegen allerlei ziekten, (e) wonden snel geneest. De uitspraken b tot en met e brengen gezamenlijk één bepaalde gedachte tot uitdrukking nl. dat bloed verschillende taken vervult. De uitspraken b tot en met e vervullen de rol van illustrerende bijzonderheden of voorbeelden van de hoofdgedachte.

Na de omschrijving van de taak en de criteria waaraan de teksten moesten voldoen, werd een aantal teksten geselecteerd en bewerkt. Evenals in het voorgaande onderzoek bleek ook nu dat de selectie van teksten een tijdrovende aangelegenheid was, omdat het merendeel van de informatieve teksten waarmee kinderen op school worden geconfronteerd, bestaat uit een opsomming van feiten zonder veel samenhang. Nadat de alternatieven waren geformuleerd, werd een uitvoerige instructie geschreven voor de afname van de eerste verzameling items.

De ontwikkeling van de test

Voor de constructie van twee parallelvormen van de hoofdgedachtetest was een grote verzameling items nodig, die voldoende was gezuiverd. In totaal vonden drie proefafnames van de eerste verzameling items plaats en twee proefafnames van de tweede verzameling items. Het resultaat van deze afnames was een verzameling van 65 items die wat betreft moeilijkheidsgraad en item-totaalcorrelatie voldeden aan de eisen (vgl. Aarnoutse, 1979).

Op basis van een inhoudelijk criterium werden de 65 items vervolgens over vier testdelen (deel A, B, C en D) verdeeld. De testdelen werden in één week in zes vierde klassen van zes verschillende lagere scholen aangenomen ($n = 186$). Om op mogelijke volgorde-effecten te controleren, werd de afname-volgorde van de testdelen systematisch gevarieerd. De resultaten waren bevredigend (zie tabel 4.1.). De test en de testdelen voldeed respectievelijk voldeden aan de

eis van betrouwbaarheid (KR-20) nl. hoger dan .80. Enkele items werden verwijderd vanwege een te hoge p-waarde ($> .85$) of vanwege een te lage itemrest-correlatie ($< .25$).

Wat betreft de constructie van de parallelvormen werden de volgende eisen geformuleerd:

- de parallelvormen dienen elk een bepaald aantal typen teksten (zes) te bevatten en een gelijk aantal teksten voor elk type;
- de parallelvormen dienen wat betreft gemiddelde, spreiding en betrouwbaarheid zo min mogelijk te verschillen; de betrouwbaarheid moet tevens maximaal zijn;
- de parallelvormen dienen zo hoog mogelijk met elkaar te correleren.

Bij de constructie van de parallelvormen werden de items in 12 groepen verdeeld, die op 32 mogelijke manieren gecombineerd konden worden tot twee parallelvormen. Van elke combinatie werden de psychometrische gegevens berekend. Bij de keuze van de definitieve combinatie kreeg de correlatie tussen de parallelvormen het grootste gewicht.

De belangrijkste gegevens van de parallelvormen worden in tabel 4.1. vermeld.

Tabel 4.1. Gegevens van de Hoofdgedachtetest. Vorm A en B.

Oorspronkelijke testgegevens	Vorm A		deel A1
	aantal items:	29	15 items
	gemid. p. :	.61	
	stand. dev. :	6.95	
	KR-20 :	.89	
	Vorm B		deel A2
	aantal items:	29	14 items
	gemid. p. :	.60	
	stand. dev. :	6.99	
	KR-20 :	.89	
			deel B1
	aantal items:	29	15 items
	gemid. p. :	.60	
	stand. dev. :	6.99	
	KR-20 :	.89	
			deel B2
	aantal items:	29	14 items
	gemid. p. :	.60	
	stand. dev. :	6.99	
	KR-20 :	.89	

De correlatie tussen de scores op vorm A en B bedroeg .92. Gezien de gevolgen van de procedure zullen de correlatie tussen de vormen en het verschil tussen de vormen in gemiddelde, standaarddeviatie en KR-20 alle enigszins geflatteerd zijn.

De vormen A en B bevatten geen items met een itemrestcorrelatie lager dan .25. In totaal waren er slechts 4 items met een itemrestcorrelatie tussen .25 en .30. Wat betreft de moeilijkheidsgraad van de items lagen de p-waarden bij beide vormen tussen .30 en .85 (on gecorrigeerd voor raden).

Uit een principale componenten-analyse op de correlaties tussen de items bleek dat zowel bij vorm A als bij vorm B sprake was van één algemene factor en dat de restfactoren van weinig betekenis waren.

Ten behoeve van de afname werd elke parallelvorm in delen gesplitst nl. deel A1 en A2 en deel B1 en B2. De items van de parallelvormen werden zo gelijk mogelijk over de delen verdeeld. Hierbij werd rekening gehouden met het aantal items per type tekst en de moeilijkheidsgraad van de items. De volgorde van de items A1 en A2 en B1 en B2 werd met behulp van een tabel met random getallen bepaald.

4.4.2. De Bedoelingstest

Taak en kenmerken van de teksten

In hoofdstuk II is er op gewezen dat het afleiden van de bedoeling van de hoofdpersoon uit een fictieve tekst de nodige kennis veronderstelt van min of meer conventionele handelingspatronen en van kennis van de structuur van verhalende teksten. Verondersteld wordt dat kinderen evenals volwassenen handelingen interpreteren als onderdelen van een plan en uit het resultaat van de uitvoering van een plan het motief van de handelingen kunnen afleiden. Bovendien wordt verondersteld dat kinderen gebruik kunnen maken van hun kennis van structuur van verhalen.

Het afleiden van de bedoeling van de hoofdpersoon uit een fictieve tekst wordt omschreven als het vinden van een handelingsmotief in een verhalende tekst waarin een handelingssequentie wordt beschreven als onderdeel van de uitvoering van een plan, zonder dat wordt aangegeven wat het doel hiervan is. Met de term "afleiden" wordt ook hier niet bedoeld dat de leerlingen het handelingsmotief van de hoofdpersoon zelf moeten genereren of formuleren. Na de tekst en een vraag die elementen bevat die letterlijk aan de tekst zijn ont-

leend, moeten de leerlingen het meest adequate motief uit een viertal alternatieven selecteren. Opgemerkt moet worden dat de vraag naar het handelingsmotief van de hoofdpersoon een vraag is naar de uiteindelijke bedoeling (motief), voor zover die uit de tekst is af te leiden. Wanneer de tekst dus aanwijzingen bevat voor een dieperliggende bedoeling, dan is deze het juiste antwoord op de vraag.

Na de omschrijving van de taak werden criteria geformuleerd waaraan de teksten moesten voldoen. Deze criteria waren:

- a. de teksten moeten een verhalend karakter hebben en niet beschrijvend of beschouwend zijn;
- b. de teksten bevatten tenminste één handelende persoon;
- c. in de tekst wordt een aantal handelingen beschreven die zich binnen een bepaalde plaats- en tijdsruimte afspelen;
- d. de tekst bevat een aantal aanwijzingen voor de bedoeling van de hoofdpersoon. Deze aanwijzingen kunnen expliciet zijn (bijvoorbeeld door het gebruik van werkwoorden als "willen" en "van plan zijn"), maar ook impliciet;
- e. de tekst bevat een resultaat van het handelen;
- f. de teksten moeten niet te veel moeilijke woorden bevatten en geen ingewikkelde zinsconstructies;
- g. de teksten moeten ontleend of bewerkt worden uit het leesmateriaal van kinderen van 8 tot 12 jaar.

Ter verduidelijking van het bovenstaande volgt nu een item uit de definitieve versie van de Bedoelingtest (Vorm A, item 13).

Een vos en een haas liepen samen over de besneeuwde velden. Het was koud want het vroor dat het kraakte. Ze hadden erge honger, maar het was moeilijk om in deze barre tijd aan eten te komen. Gelukkig zagen ze in de verte een meisje dichterbij komen. Ze had een mandje aan haar arm en uit dat mandje steeg een geur van heerlijke, verse broodjes op. "Weet je wat?" zei de vos tegen de haas. "Als jij nu languit op de grond gaat liggen en je niet beweegt, dan zet dat meisje vast haar mandje neer om je op te pakken. Want uit jouw pels kan ze warme handschoenen maken... Dan pak ik haar mandje weg!" De haas deed wat de vos gezegd had, terwijl de vos achter een sneeuwhoop verdween. Het meisje kwam langs en zag het zielige haasje liggen. Ze zette haar mandje neer en boog zich over de

haas. Meteen kwam de vos tevoorschijn. Hij greep het mandje en schoot het veld in. De haas sprong vlug overeind en holde de vos achterna. Het leek alsof de vos hem kwijt wilde raken. Toen ze langs een visvijver kwamen, riep de haas "zullen we verse vis bij de broodjes nemen?". De vos stopte, want daar had hij wel zin in. "Laat je staart in het water hangen, dan zullen de vissen er beslist in bijten", zei de haas. De vos gehoorzaamde meteen. Maar het duurde niet lang of de staart zat vastgevroren in het ijs. Ondertussen at de haas de broodjes één voor één smakelijk op. Daarna zei hij tegen de vos: "Wacht maar tot de dooi invalt."

13. Waarom zei de haas "Zullen we verse vis bij de broodjes nemen"?

Dat zei de haas....

1. omdat hij wist dat een gat in het ijs vlug dicht zou vriezen.
2. omdat hij de hebberige vos wilde straffen.
3. omdat hij wist dat de vos met zijn staart vissen kon vangen.
4. omdat hij wist dat de vos dat ook lekker vond.

Het is mogelijk de structuur van dit verhaal waarin twee hoofdpersonen en twee episodes voorkomen aan de hand van het verhaalschema van Rumelhart (1977b) weer te geven (vgl. Aarnoutse et al., 1979).

Zoals uit het voorbeelditem blijkt, is een letterlijk gedeelte uit de tekst in de vraag verwerkt. Dit gedeelte heeft steeds betrekking op een belangrijk aspect van de uitvoering van het plan, dat de hoofdpersoon (protagonist) uitvoert als middel om zijn doel te bereiken. Het is van belang dat de leerlingen niet het doel van het plan als bedoeling geven (alternatief 1). Dit is slechts één mogelijke manier om het uiteindelijke doel (alternatief 2) te bereiken.

Na de omschrijving van de taak en de criteria waaraan de teksten moesten voldoen, werd een aantal teksten geselecteerd en bewerkt. Bovendien werd een uitvoerige instructie geschreven voor de afname van de eerste verzameling items.

De ontwikkeling van de test

Voor de constructie van de parallelvormen van de bedoelingstest was eveneens een grote verzameling items nodig, die voldoende was gezuiverd. In to-

taal vonden vier proefafnames van de eerste verzameling items plaats en één proefafname van de tweede verzameling items. Bij de proefafnames van de eerste verzameling items werd voor de constructie van de antwoordalternatieven aanvankelijk gebruik gemaakt van de antwoorden van de leerlingen. De testvragen werden hiertoe als open vragen gepresenteerd. Bij latere proefafnames werden de antwoordalternatieven rechtstreeks geconstrueerd zonder gebruik te maken van antwoorden van leerlingen op open vragen. Het resultaat van deze afnames was een verzameling van 56 items die wat betreft moeilijkheidsgraad en itemrestcorrelatie niet alle voldeden aan de eisen. Om die reden werd deze verzameling in haar geheel nog eens afgenomen. Daarna volgde nog een afname van een combinatie van een aantal items. Het resultaat was een verzameling van 56 items die wat betreft moeilijkheidsgraad en itemrestcorrelatie grotendeels voldeden aan de eisen ($.25 < p < .85$; $r_{ir} > .25$).

Voor de afname ten behoeve van de constructie van de parallelvormen werden de 56 items in 4 delen (deel A, B, C en D) verdeeld. De testdelen werden binnen een tijdsbestek van twee weken afgenomen in vijf vierde klassen van vier verschillende scholen ($n = 127$). Op mogelijke volgorde-effecten werd gecontroleerd door de volgorde van afname van de vier testdelen systematisch te variëren. De resultaten waren niet geheel bevredigend. Van de 56 items voldeden er 16 niet aan de gestelde eisen. Van deze 16 items werden 8 items geklassificeerd als grensgevallen; deze items hadden een p-waarde (ongecorrigeerd voor raden) tussen .85 en .88 en/of een itemrestcorrelatie tussen .20 en .25. Besloten werd deze grensgevallen te handhaven, daar ze bij andere afnames wel voldeden aan de gestelde eisen. In totaal bleven er dus 48 items over voor de constructie van de parallelvormen.

Voor de constructie van de parallelvormen werden de volgende eisen geformuleerd:

- a. de parallelvormen moeten wat de inhoud van de teksten betreft zo gelijk mogelijk zijn;
- b. de parallelvormen dienen wat betreft gemiddelde, spreiding en betrouwbaarheid zo min mogelijk te verschillen; de betrouwbaarheid moet tevens maximaal zijn;
- c. de parallelvormen dienen zo hoog mogelijk met elkaar te correleren.

Bij de constructie van de parallelvormen werden de items in twee groepen verdeeld, die op twee verschillende manieren gecombineerd konden worden tot twee

parallelvormen. Van elke combinatie werden de psychometrische gegevens berekend. Bij de keuze van de definitieve combinatie kreeg de correlatie tussen de parallelvormen het grootste gewicht.

De belangrijkste gegevens van de parallelvormen worden in tabel 4.2. vermeld.

Tabel 4.2. Gegevens van de Bedoelingtest. Vorm A en B.

Oorspronkelijke testgegevens	Vorm A		deel A1
	aantal items:	24	12 items
	gemid. p. :	.68	
	stand. dev. :	4.8	deel A2
	KR-20 :	.83	12 items
	Vorm B		deel B1
aantal lln. :	aantal items:	24	12 items
aantal items:	gemid. p. :	.68	
gemid. p. :	stand. dev. :	4.7	deel B2
stand. dev. :	KR-20 :	.82	12 items
KR-20 :			

De correlatie tussen de scores op vorm A en B bedroeg .85. Gezien de gevolgsprocedure zullen de correlatie tussen de parallelvormen en het verschil in gemiddelde, standaarddeviatie en KR-20 alle enigszins geflatteerd zijn.

Uit een principale componenten-analyse op de correlaties tussen de items bleek, dat zowel bij vorm A als bij vorm B sprake was van één algemene factor en dat de restfactoren van weinig betekenis waren.

Ten behoeve van de afname werd elke parallelvorm in twee delen gesplitst nl. deel A1 en A2 en deel B1 en B2. De items van de parallelvormen werden zo gelijk mogelijk over de delen verdeeld. Hierbij werd rekening gehouden met de inhoud van de items (Fabels versus Rest) en de moeilijkheidsgraad van de items. De volgorde van de items in een deel werd bepaald met behulp van een tabel met randomgetallen.

4.4.3. De Redeneertest

Taak en kenmerken van de teksten

Zoals uit hoofdstuk II blijkt, komt slechts een beperkt aantal redeneervormen in aanmerking om te worden opgenomen in een redeneertest bedoeld voor leerlingen van het vierde leerjaar van de lagere school. Met name de geldige redeneervormen *modus ponens* en *modus tollens* en de ongeldige redeneervormen *conversie* en *inversie* lijken geschikt.

In het volgende wordt de redeneertaak vooral aan de hand van voorbeelden verduidelijkt. Om een indruk te geven van de wijze waarop deze taak is geconcretiseerd, volgt eerst een item uit de definitieve versie van de Redeneertest (vorm B, item 28).

Mijnheer Peters is een vriendelijke man die bij ons in de buurt woont. Hij heeft een oud autootje dat hij met veel zorg onderhoudt. Soms wil de auto van Peters niet meteen starten. Je ziet dan dat hij zich ergert en van alles doet om het oudje weer aan de praat te krijgen. Als de auto meteen start, steekt Peters altijd een dikke sigaar op. Peters rookt wel vaker dikke sigaren, maar als hij tevreden is smaken ze hem beter. Gisteren stak Peters een dikke sigaar op.

28. Startte de auto van Peters gisteren meteen?

1. Er staat te weinig in het verhaal om deze vraag te kunnen beantwoorden.
2. Ja: de auto van Peters startte gisteren meteen.
3. Nee: de auto van Peters startte gisteren niet meteen.

De taak bestaat uit de opdracht een conclusie-vraag te beantwoorden over een tekst die een major en een minor bevat.

Elk item van de test bevat een major, die categorisch of hypothetisch is. Een voorbeeld van een categorische major is: Alle schriften van Karel zijn groen. Een voorbeeld van een hypothetische major is: Als de auto meteen start, steekt Peters altijd een dikke sigaar op. Een major bestaat uit twee termen: A en B. In het eerste voorbeeld heeft de A-term betrekking op "schriften van Karel" en de B-term op "groene schriften". In het tweede voor-

beeld heeft de A-term betrekking op "keren dat Peters de auto start" en de B-term op "keren dat Peters een dikke sigaar opsteekt".

Naast een major bevat elk item van de test een minor. Een minor heeft betrekking op één persoon, object of gebeurtenis. Er zijn vier typen minors: de minor kan de A-term van de major bevestigen (+) of ontkennen (-), of de B-term van de major bevestigen of ontkennen. Een voorbeeld is:

Alle schriften van Karel zijn groen (categorische major)

Dit schrift is van Karel (minor + A)

Dit schrift is niet van Karel (minor - A)

Dit schrift is groen (minor + B)

Dit schrift is niet groen (minor - B)

Het voorbeelditem bevat een minor van het type + B: "Gisteren stak Peters een dikke sigaar op".

Bij elke tekst van een item wordt een conclusie-vraag gesteld. Deze vraag heeft betrekking op de conclusie die uit de major en de minor getrokken kan worden.

Een voorbeeld is: Alle schriften van Karel zijn groen (major)

Dit schrift is van Karel	(minor)

Dit schrift is groen	(conclusie)

De conclusie "Dit schrift is groen" wordt getransformeerd tot de conclusie-vraag "Is dit schrift groen?". De conclusie-vraag wordt steeds in positieve termen geformuleerd. Dit betekent dat er géén vragen worden gesteld als: "Is dit schrift niet groen?". Een conclusie-vraag met een ontkenning zou tot ongewenste dubbele ontkenningen leiden, omdat ook het antwoord een ontkenning kan bevatten.

Bij elke major zijn twee conclusie-vragen mogelijk. Als de minor betrekking heeft op de A-term van de major (+A of -A), vraagt de conclusie-vraag naar de B-term van de major. Als de minor betrekking heeft op de B-term van de major (+B of -B), vraagt de conclusie-vraag naar de A-term van de major. Een voorbeeld is: Major: Alle schriften van Karel zijn groen.

Minor: Dit schrift is van Karel (+A), of: Dit schrift
is niet van Karel (-A).

Conclusie-vraag: Is dit schrift groen? (B?).

In het voorbeelditem luidt de conclusie-vraag: "Startte de auto van Peters

gisteren meteen?" Deze vraag is een vraag naar de A-term van de major, want de minor heeft betrekking op de B-term.

Het juiste antwoord op de conclusie-vraag wordt geheel bepaald door de minor. Of een major categorisch of hypothetisch is, is in dit verband niet belangrijk. Als de minor van het type + A is, moet de conclusie-vraag met "Ja" beantwoord worden. Bij een minor - B is het juiste antwoord "Nee". Bij een minor - A of een minor + B is het antwoord op de conclusie-vraag "Onzeker". De leerlingen moeten dan als antwoord geven: "Er staat te weinig in het verhaal om deze vraag te kunnen beantwoorden". In het voorbeelditem is de minor van het type + B. Het antwoord op de conclusie-vraag is dus "Onzeker". In de test worden bij elke vraag drie dezelfde antwoordalternatieven gegeven nl.

1. Er staat te weinig in dit verhaal om deze vraag te kunnen beantwoorden.
2. Ja:
3. Nee:

Uit het voorgaande blijkt dat er twee typen majors zijn nl. categorische (C) en hypothetische (H) en vier typen minors nl. + A, - A, + B, - B. Een combinatie van twee typen majors en vier typen minors resulteert in acht typen redeneringen. In tabel 4.3. wordt elk type redenering aangeduid door het major- en minortype te vermelden. Een categorische major met een minor van het type + A geeft een redenering van het type C + A. In de onderstaande tabel zijn de acht typen redeneringen schematisch weergegeven.

Tabel 4.3. Typen redeneringen

		major	
		C	H
minor	+ A	C + A	H + A
	- A	C - A	H - A
	+ B	C + B	H + B
	- B	C - B	H - B

Het voorbeelditem heeft een hypothetische major en een minor van het type + B. Het bevat dus een redenering van het type H + B. Opgemerkt moet worden dat bij de constructie van de Redeneertest rekening is gehouden met het aantal items per type redenering.

Om de conclusie-vraag te kunnen beantwoorden, hebben de leerlingen strikt logisch gezien alleen de major en de minor nodig. In sommige gevallen is echter in de tekst een B > A-zin opgenomen. Deze zin geeft aan dat de B-term van de major groter is dan de A-term. De functie van de zin is een onjuiste interpretatie van de major te voorkomen. Een voorbeeld van een B > A-zin is: Major: Als opa De Vries zijn krantje leest (A), heeft hij altijd zijn bril op (B).

B > A-zin: Opa heeft zijn bril ook wel eens op als hij zijn krantje niet leest.

Het voorbeelditem bevat ook een B > A-zin. Bij de major "Als de auto meteen start, steekt Peters altijd een dikke sigaar op" wordt de B > A-zin gegeven "Peters rookt wel vaker dikke sigaren,.....".

Binnen een major wordt geen uitspraak gedaan over de omvang van de A- en de B-term. Nu is het voor conclusie-vragen waarop het juiste antwoord "onze-ker" is echter van groot belang dat de leerlingen de major niet interpreteren als A = B (A valt samen met B, A is even groot als B). Als de leerlingen de major betreffende opa De Vries interpreteren als "Als opa De Vries zijn krantje leest, heeft hij altijd zijn bril op, en anders nooit", zullen ze de conclusie-vraag beantwoorden met "Nee". Om deze interpretatie tegen te gaan, wordt een B > A-zin toegevoegd. Opgemerkt moet worden dat in de test alleen een B > A-zin wordt toegevoegd in die gevallen, waarin het juiste antwoord op de conclusie-vraag "onze-ker" is, en waarin bovendien niet uit elementaire kennis duidelijk is, dat de B-term van de major groter is dan de A-term.

Besloten is hypothetische majors de vorm te geven van "Als A, dan altijd B". Aan alle hypothetische majors wordt het woord "altijd" toegevoegd om misverstanden over de betekenis ervan te voorkomen. De toevoeging maakt expliciet wat in feite bedoeld wordt.

Naast de major, de minor en soms een B > A-zin, bestaat een tekst nog uit een aantal andere zinnen. Deze zijn voor de redenering niet van belang en worden daarom irrelevante zinnen genoemd. Door de irrelevante zinnen krijgen

de relevante zinnen een context. Er ontstaat een verhaaltje, een tekst.

De tekst wordt steeds zo geconstrueerd, dat niet is na te gaan of hetgeen in de tekst beschreven is empirisch waar is of niet. Een juist antwoord moet uitsluitend gegeven worden op basis van een redenering en niet op grond van kennis van de wereld. Er wordt naar gestreefd de teksten te laten handelen over onderwerpen uit de belevingssfeer van kinderen uit het vierde leerjaar (school, sport, spel en dergelijke).

Na de omschrijving van de taak en enkele criteria waaraan de teksten moesten voldoen, werd een aantal teksten geschreven. Bovendien werd een uitvoerige instructie opgesteld voor de afname van de eerste verzameling items.

De ontwikkeling van de test

Voor de constructie van twee parallelvormen van de redeneertest was evenals bij de andere taakgerichte tests een grote verzameling items nodig, die voldoende was gezuiverd. In totaal vonden zes proefafnames van de eerste verzameling items plaats en twee proefafnames van de tweede verzameling. Uit de eerste proefafname (van de eerste verzameling) bleek dat de betrouwbaarheid (KR-20) van de test erg laag was (.58). Langs twee wegen werd geprobeerd een oplossing te vinden voor dit probleem. In de eerste plaats werd onderzocht of de antwoordalternatieven voor verbetering vatbaar waren. De eerste vijf proefafnames hadden vooral tot doel om geschikte antwoordalternatieven te vinden; bij deze proefafnames werden steeds verschillende alternatieven gebruikt. In de tweede plaats werd veel aandacht besteed aan de vraag of de verzameling items verdeeld moest worden in subtests. Uit een principale componenten-analyse op de itemcorrelaties tussen subtests en uit de KR-20 coëfficiënten bleek bij verschillende proefafnames, dat de items waarop het juiste antwoord "ja" (de typen redeneringen $C + A$ en $H + A$) of "nee" (de typen redeneringen $C - B$ en $H - B$) was, en de items waarop het juiste antwoord "onzeker" (de typen redeneringen $C - A$, $H - A$, $C + B$ en $H + B$) was, twee aparte groepen van items vormden. Uit verdere analyses bleek dat deze subtestvorming voor een deel was te verklaren door de relatief grote verschillen tussen leerlingen in de frequenties waarmee antwoorden van een bepaald type (ja, nee of onzeker) gegeven werden. Getracht is correcties te ontwerpen voor verschillen in antwoordfrequenties. Het gebruik van gecorrigeerde subtestscores bleek verdedigbaar, maar gecorrigeerde itemscores waren weinig zinvol, hetgeen bleek uit een simulatieproef met randomscores.

Voor de vijfde proefafname werd besloten de antwoordalternatieven te kiezen zoals weergegeven in het voorbeelditem. Verder werd besloten een B > A-zin in de betreffende teksten op te nemen en het woord "altijd" aan alle hypothetische majors toe te voegen. Tenslotte werden subtests onderscheiden. De items waarop het juiste antwoord "ja" of "nee" was, vormden de subtest "Zeker" en de items waarop het juiste antwoord "onzeker" was ("Er staat te weinig in het verhaal om deze vraag te kunnen beantwoorden") vormden de subtest "Onzeker". De resultaten van deze proefafname en de volgende afnames waren positief. In totaal kwamen op deze wijze 80 items beschikbaar, die voldeden aan de eisen wat betreft moeilijkheidsgraad en itemrestcorrelatie. Aan deze verzameling werden nog 12 reserve-items toegevoegd.

Voor de afname ten behoeve van de constructie van de parallelvormen werden 92 items in 4 delen (deel A, B, C en D) verdeeld, die in twee weken in zes vierde klassen van zes verschillende lagere scholen werden afgenomen ($n = 161$). Op mogelijke volgorde-effecten werd gecontroleerd door de volgorde van de afname van de vier testdelen systematisch te variëren. De resultaten van deze afname waren bevredigend. Vervolgens werd het aantal items gereduceerd tot 80 waarvan 48 items voor de subtest Zeker en 32 items voor de subtest Onzeker. De belangrijkste gegevens van deze verzameling items worden vermeld in tabel 4.4.

Wat betreft de constructie van de parallelvormen werden de volgende eisen geformuleerd:

- a. er dienen parallelvormen van de subtest Zeker en Onzeker gevormd te worden. De vier logische typen van de subtest Zeker ($C + A$, $H + A$, $C - B$, $H - B$) bevatten elk 6 items; de vier logische typen van de subtest Onzeker ($C - A$, $H - A$, $C + B$, $H + B$) bevatten elk 4 items;
- b. de parallelvormen dienen wat betreft gemiddelde, spreiding en betrouwbaarheid zo min mogelijk te verschillen; de betrouwbaarheid moet tevens maximaal zijn;
- c. de parallelvormen dienen zo hoog mogelijk met elkaar te correleren.

Bij de constructie van de parallelvormen werden de items van zowel de subtest Zeker als Onzeker in 8 groepen verdeeld, die op 8 mogelijke manieren gecombineerd konden worden tot twee parallelvormen. Van elke combinatie werden de psychometrische gegevens berekend. Bij de keuze van de definitieve combinatie kreeg de correlatie tussen de parallelvormen het grootste gewicht.

Tabel 4.4. Gegevens van de Redeneertest. Vorm A en B.

	C	H		Totaal	Zeker	Onzeker
+ A	12	12	items	80	48	32
- A	8	8	\bar{p}	.65	.72	.53
+ B	8	8	s	14.5	9.2	8.5
- B	12	12	KR-20	.93	.91	.92

\bar{p} = gemiddelde moeilijkheidsgraad (niet gecorrigeerd voor raden)
 s = standaarddeviatie
 C = categorische major
 H = hypothetische major
 +A = minor die A-term bevestigt
 -A = minor die A-term ontkent
 +B = minor die B-term bevestigt
 -B = minor die B-term ontkent

		Vorm A				
		C	H	Totaal Zeker Onzeker		
+ A	6	6	items	40	24	16
- A	4	4	\bar{p}	.65	.73	.53
+ B	4	4	s	7.5	4.8	4.5
- B	6	6	KR-20	.87	.84	.85

A Totaal x B Totaal : .90
 A Zeker x B Zeker : .89
 A Onzeker x B Onzeker : .88

Vorm B			Totaal	Zeker	Onzeker	
	C	H				
+ A	6	6	items	40	24	16
- A	4	4	\bar{p}	.65	.72	.53
+ B	4	4	s	7.4	4.7	4.4
- B	6	6	KR-20	.87	.83	.85

deel A1	C	H
+ A	3	3
- A	2	2
+ B	2	2
- B	3	3

deel A2	C	H
+ A	3	3
- A	2	2
+ B	2	2
- B	3	3

deel B1	C	H
+ A	3	3
- A	2	2
+ B	2	2
- B	3	3

deel B2	C	H
+ A	3	3
- A	2	2
+ B	2	2
- B	3	3

De belangrijkste gegevens van de parallelvormen worden in tabel 4.4. vermeld.

De correlatie tussen de scores op vorm A en B bedroeg .90. Gezien de gevolgtrede procedure zullen de correlaties tussen de parallelvormen en het verschil tussen de vormen in gemiddelde, standaarddeviatie en KR-20 alle enigszins geplateerd zijn.

Bij de itemanalyse bleek dat er drie items zijn met een itemrestcorrelatie lager dan .25 (laagste correlatie .20) en vier items met een p-waarde groter dan .85 (niet gecorrigeerd voor raden).

Evenals in de proefafnames treden ook in de parallelvormen de Zekere en Onzekere items als aparte subtests naar voren. De correlatie tussen alle Zekere items (48) en alle Onzekere items (32) is slechts .34. Uit een principale componentenanalyse op de items van vorm A respectievelijk B blijkt duidelijk dat er sprake is van twee factoren. De factorstructuur van beide vormen is gelijk. Factor 1 kan geïnterpreteerd worden als Onzeker, factor 2 als Zeker. Ten behoeve van de afname werd elke parallelvorm in delen gesplitst nl. deel A1 en A2 en deel B1 en B2. Hierbij werd rekening gehouden met het aantal items per type redenering en de moeilijkheidsgraad van de items. De volgorde van de items in een deel werd bepaald met behulp van een tabel met random getallen.

4.4.4. De Free Response Cloze Test

Achtergrond

In hoofdstuk III is een overzicht gegeven van de onderzoeksliteratuur met betrekking tot de cloze procedure. Het kenmerkende van de standaard of free response cloze procedure bestaat hierin dat woorden op een systematische wijze uit een tekst worden weggelaten en proefpersonen wordt gevraagd de ontbrekende woorden in te vullen. Bij deze procedure wordt elk n-de woord weggelaten, waarbij voor n meestal een getal tussen de vijf en de tien wordt gekozen.

Als voordeel van deze procedure wordt onder andere genoemd, dat een free response cloze test gemakkelijk en in korte tijd is te construeren. Bovendien wordt volgens verschillende auteurs een meer objectieve en meer uitgebreide steekproef getrokken uit de syntactische en semantische inhoud van een tekst dan met de gangbare testprocedures het geval is. Verder wordt als

voordeel genoemd dat het proces van begrijpend lezen niet wordt verstoord door vragen die soms moeilijker te begrijpen zijn dan de tekst waarop ze betrekking hebben.

Een nadeel van de free response cloze test kan zijn, dat te veel de nadruk wordt gelegd op de syntactische aanwijzingen in de tekst. Een ander bezwaar kan zijn dat minimaal 50 woorden uit een tekst moeten worden weggelaten. Wanneer minder woorden worden weggelaten, neemt de kans toe dat toevalsfouten de representativiteit van de steekproef van woorden aantasten. Een volgend nadeel is dat de test sterk afhankelijk is van de moeilijkheidsgraad en het type tekst dat gebruikt wordt. Om dit bezwaar op te heffen kunnen meerdere teksten worden gebruikt. Een laatste nadeel is dat de test met de hand moet worden nagekeken wat bij grote groepen leerlingen een tijdsintensieve aangelegenheid wordt.

Bij het nakijken van een free response cloze test kunnen, zoals we gezien hebben, verschillende maatstaven worden aangelegd voor wat onder een juist antwoord en wat onder een foutief antwoord dient te worden verstaan. De meest gebruikte scoringswijze is die waarbij men alleen het woord dat oorspronkelijk in de tekst stond, als een goed antwoord klassificeert, de zogenaamde exacte woord scoringswijze. Anderen geven er de voorkeur aan om ook synoniemen en andere woorden die syntactisch en semantisch in de context passen, juist te noemen. De laatste scoringswijze wordt wel de semantisch juiste scoring genoemd. Door voor het oorspronkelijke woord een andere code te gebruiken dan voor de antwoorden die semantisch juist zijn, is het niet noodzakelijk dat er bij de codering al een beslissing wordt genomen omtrent het gebruik van een bepaalde scoringswijze bij de uiteindelijke analyse. Dit kan men eventueel laten afhangen van de nakijk-consistentie van de semantisch juiste scoringswijze.

De ontwikkeling van de test

Voor de constructie van twee parallelvormen van de Free Response Cloze Test werden vier informatieve teksten geselecteerd uit het leesmateriaal bestemd voor leerlingen van het vierde en vijfde leerjaar van de lagere school. Voor deze teksten werden twee leesbaarheidsindices berekend nl. de leesindex A van Brouwer (1963) en de CF4 van Zondervan, Van Stam en Gunneweg (1976). Op basis van de leesindex A konden de teksten worden gekarakteriseerd als ge-

makkelijk tot zeer gemakkelijk. Op grond van de CF4 kon een foutenpercenta-ge van 40 tot 50% worden verwacht.

De vier teksten waren alle ongeveer 400 woorden lang. Elke tekst had een inleidende passage van ongeveer 50 woorden. Daarna werd elk zevende woord weggelaten. In elke tekst werden 50 woorden weggelaten.

Na de selectie van de teksten werd een instructie geschreven voor de afname van de vier testvormen. In de instructie wordt er op gewezen dat eerst het hele verhaal moet worden gelezen alvorens met het invullen van de open plaatsen te beginnen. Bovendien wordt geadviseerd niet te lang bij een open plaats stil te blijven staan en deze eventueel later in te vullen.

De vier testvormen werden in vier klassen van twee scholen afgenomen ($n = \pm 100$). Om mogelijke volgorde-effecten zo veel mogelijk te spreiden, werd de volgorde van afname van de vier testvormen zodanig gevarieerd dat alle permutaties van mogelijke afnamevolgorden nagenoeg even vaak voorkwamen. Bij het nakijken werden twee scoringswijzen toegepast nl. de exacte woord scoringswijze en de semantisch juiste scoringswijze. De gemiddelde moeilijkheidsgraad van de vier testvormen lag bij de exacte woord scoringswijze rond hetzelfde niveau nl. tussen .51 en .55 hetgeen opvallend goed overeenstemt met het voorspelde percentage cloze fouten berekend met de CF4 formule van Zondervan c.s. Bij de semantisch juiste scoringswijze lag de gemiddelde moeilijkheidsgraad bij drie testvormen ongeveer .10 hoger.

De KR-20 betrouwbaarheidscoëfficiënten van de vier testvormen bedroegen bij de exacte woord scoringswijze .74, .74, .79 en .71 en bij de semantisch juiste scoringswijze .77, .79, .85 en .77. De waarden van deze coëfficiënten zijn bij de semantisch juiste scoringswijze steeds hoger dan bij de exacte woord scoringswijze. Deze waarden zijn nog acceptabel voor vergelijkingen tussen groepen maar te laag voor een interpretatie per individu.

Wat betreft de samenstelling van twee parallelvormen moest een keuze worden gedaan uit de vier testvormen. Deze keuze werd voornamelijk gemaakt op grond van de psychometrische gegevens. Daarbij speelde de onderlinge correlatie een belangrijke rol, terwijl er ook rekening mee werd gehouden dat de waarden van de KR-20, de gemiddelde p-waarde en de spreiding van de twee te kiezen vormen zo min mogelijk van elkaar verschilden. De belangrijkste gegevens van de parallelvormen worden in tabel 4.5. vermeld.

Tabel 4.5. Gegevens van de Free Response Cloze Test. Vorm A en B.

	FRCT-A		FRCT-B	
	exacte woord- scoring	semantisch juiste scoring	exacte woord- scoring	semantisch juiste scoring
aantal lln.	103	103	98	98
aantal items	50	50	50	50
gemid. score	26.1	31.1	26.2	32.4
stand. dev.	5.25	5.96	5.24	5.93
gemid. p.	.52	.62	.52	.65
KR-20	.74	.79	.71	.77

De onderlinge correlatie tussen de scores op vorm A en B was voor de exacte scoringswijze .65 en voor de semantische scoringswijze .74. De correlaties tussen de beide scoringswijzen per test zijn beide .95. De overeenstemming tussen twee onafhankelijke beoordelaars, berekend bij de FRCT-vorm B was bij de semantisch juiste scoringswijze 92.9% en bij de exacte woord scoringswijze 98.2%. Om de consistentie van de semantisch juiste scoringswijze te verhogen werd besloten om per item aan te geven welke woorden wel en welke beslist niet als semantisch juist aangemerkt dienen te worden. Voor de leerlingen werd deze test de invultest genoemd.

4.4.5. De Multiple Choice Cloze Test

Achtergrond

In hoofdstuk III is de nodige aandacht besteed aan de multiple choice cloze tests zoals die door O'Reilly, Schuder en Kidder (1976) zijn ontwikkeld en aan het validiteitsonderzoek dat in verband met deze tests is uitgevoerd.

In dit onderzoek zal worden nagegaan in hoeverre een multiple choice cloze test een betrouwbare en valide maat is voor het begrijpend lezen in het algemeen en niet zoals O'Reilly c.s. claimen een maat voor het letterlijk begrijpen van lezen. Bij de ontwikkeling van deze test zal zo veel mogelijk worden uitgegaan van het ontwerp en de procedure die O'Reilly c.s.

voorstellen. Dit betekent onder andere dat alleen inhoudswoorden worden weggelaten en de afleiders zodanig worden gekozen dat deze grammaticaal juist zijn, maar semantisch niet in de tekst passen.

Bij de multiple choice cloze test wordt voor elk weggelaten woord in de tekst een aantal alternatieven aangeboden waarvan er telkens één moet worden gekozen. Hiermee wordt een belangrijk nadeel van de free response cloze procedure opgeheven nl. het tijdrovende nakijkwerk bij grote groepen proefpersonen. De eenvoud waarmee een free response cloze test kan worden geconstrueerd, is daarmee echter eveneens verdwenen.

De ontwikkeling van de test

Zoals vermeld, is de multiple choice cloze test voor een deel geconstrueerd volgens het ontwerp en de procedure van O'Reilly c.s. Dit betreft met name de aard en de omvang van de teksten, het aantal antwoordalternatieven, de keuze van de alternatieven en de lay-out.

Voor de constructie van twee parallelvormen werden in totaal 20 teksten geselecteerd. Bij de keuze van de teksten werden de volgende criteria gehanteerd:

- a. de tekst is een min of meer afgeronde informatie-eenheid met een lengte van minimaal 70 en maximaal 85 woorden;
- b. de tekst is afkomstig uit leesmateriaal bedoeld voor leerlingen van het derde tot en met het zesde leerjaar;
- c. de tekst is informatief van aard, bevat geen metaforisch taalgebruik en geen dialoog.

Om te voorkomen dat leerlingen bij de proefafname als eerste een tekst zouden krijgen die moeilijk voor hen was, werden de teksten van het eerste deel (10 teksten) en het tweede deel (10 teksten) naar moeilijkheidsgraad geordend. Daartoe werden in eerste instantie de leesbaarheidsformules van Brouwer (1963) en van Zondervan, Van Stam en Gunneweg (1976) toegepast. De omvang van de teksten was echter te beperkt om hiermee een betrouwbare ordening van teksten te verkrijgen. Om die reden werd aan tien leerkrachten van het lager onderwijs gevraagd de teksten te rangschikken naar moeilijkheidsgraad voor het vierde leerjaar. De gemiddelde intercorrelatie tussen deze individuele beoordelaars bereikte slechts een waarde van .65.

Voor het weglaten van woorden werden de volgende regels opgesteld:

- a. bepaal de woorden die in aanmerking komen om weggelaten te worden, i.c. zelfstandige naamwoorden, werkwoorden, bijvoeglijke naamwoorden en bijwoorden;
- b. na het vijfde woord wordt het eerste woord dat tot de geselecteerde woordsoorten behoort, weggelaten;
- c. het volgende woord dat wordt weggelaten, staat vijf woorden verder, mits dat eveneens tot de geselecteerde woordsoorten behoort;
- d. wanneer dat niet het geval is, wordt een woord gezocht dat wel aan de regels voldoet in de volgende volgorde: het vierde woord, het zesde, zevende, achtste etc. Minimaal bevinden zich dus minstens drie woorden tussen twee weggelaten woorden;
- e. deze procedure wordt zo lang herhaald dat 12 woorden per tekst zijn weggelaten.

Voor de keuze van de alternatieven werd gebruik gemaakt van de woordfrequentielijst van Uit den Boogaart (1975). Bij elk weggelaten woord werd de in deze publicatie gehanteerde grammaticale code aangegeven. Voor de selectie van de alternatieven werden de volgende criteria geformuleerd:

- a. alle alternatieven dienen dezelfde grammaticale code te hebben als het weggelaten woord;
- b. alle alternatieven dienen met ongeveer dezelfde frequentie voor te komen als het weggelaten woord;
- c. alle alternatieven bij een woord uit de groep enkelvoudige substantieven dienen hetzelfde lidwoord te hebben.

Voor de keuze van de alternatieven werd een zoekstrategie opgesteld binnen de genoemde woordfrequentielijst. Van deze strategie behoefde slechts enkele malen te worden afgeweken.

Voor de afname ten behoeve van de constructie van de parallelvormen werden de teksten in twee delen verdeeld en geordend naar moeilijkheidsgraad. Elke tekst kreeg eenzelfde lay-out. Ter illustratie volgt een tekst (subtest) uit de definitieve versie van de Multiple Choice Cloze Test (Vorm A, item 5).

Tenslotte werd een instructie geschreven voor de afname van de twee delen. In de instructie wordt er op gewezen dat het overzien van de tekst de keuze van het juiste alternatief kan vergemakkelijken. De leerlingen zijn echter niet verplicht de tekst eerst door te lezen alvorens een alternatief

Het postkantoor

Behalve het stadhuis staan er ----1---- andere grote gebouwen in onze ----2----. Het postkantoor is er een van. Door een ----3---- deur kom je in de ----4----. Die is hoog en ----5----. Je ziet daar een ----6---- rij loketten. Maar van de ----7---- zie je niet veel. ----8---- niet? Wel, die brieven ----9---- niet op het ----10----. Na elke buslichting worden de ----11---- op het postkantoor ----12---- en dan gaan ze, zo spoedig mogelijk, in grote postzakken naar de trein.

1 1 al
2 nog
3 ook
4 weer
5 wel

5 1 direct
2 hard
3 moeilijk
4 ruim
5 werkelijk

9 1 beven
2 blijven
3 domineren
4 klinken
5 overheersen

2 1 grond
2 manier
3 minister
4 politie
5 stad

6 1 goede
2 hele
3 kleine
4 nieuwe
5 oude

10 1 gewest
2 kistje
3 moedertje
4 postkantoor
5 vrouwtje

3 1 grote
2 hele
3 kleine
4 nieuwe
5 oude

7 1 brieven
2 negers
3 nummers
4 onderdelen
5 volken

11 1 afwijkingen
2 brieven
3 onderhandelingen
4 voorbeelden
5 zakken

4 1 gedachte
2 glimlach
3 hal
4 hoop
5 kerel

8 1 Hoe
2 Waar
3 Waarom
4 Wanneer
5 Wie

12 1 buitgemaakt
2 gestempeld
3 gewend
4 klaargelegd
5 hervat

--	--	--	--

--	--	--	--

--	--	--	--

aan te kruisen.

De twee delen werden in vier klassen van twee lagere scholen afgenomen ($n \approx \pm 100$). Op mogelijke volgorde-effecten werd gecontroleerd door de volgorde van de afname systematisch te variëren. De resultaten waren bevredigend. De gemiddelde moeilijkheidsgraad van deel I (120 items) was .77 en van deel II (120 items) .83. De KR-20 betrouwbaarheidscoëfficiënten bedroegen voor deel I .90 en voor deel II .88. De alfa-coëfficiënten, gebaseerd op de totaalscores van de 10 subtests per deel hadden ook een acceptabele waarde nl. .87 voor deel I en .82 voor deel II. Uit de gemiddelde moeilijkheidsgraad van de subtests per deel bleek dat er van een opklimming in moeilijkheid nauwelijks sprake was.

Wat betreft de samenstelling van twee parallelvormen is gewerkt met subtests als eenheden. De maximale testlengte werd vastgesteld op 8 subtests per vorm (96 items). Bij de verdeling van de 16 subtests over de beide vormen werd rekening gehouden met de inhoud van de teksten van de subtests en met de psychometrische gegevens van de testvormen die bij een bepaalde verdeling van subtests ontstonden. Op grond van de hoogte van de onderlinge correlatiecoëfficiënten van beide vormen en de overeenkomst in gemiddelde en standaarddeviatie werd die combinatie van subtests gekozen waarvan de belangrijkste gegevens in tabel 4.6. worden vermeld.

Tabel 4.6. Gegevens van de Multiple Choice Cloze Test. Vorm A en B.

	MCCT-A	MCCT-B
aantal lln.	97	97
aantal items/subtests	96/8	96/8
gemid. score	74.4	76.2
stand. dev.	9.29	9.11
gemid. p.	.78	.79
KR-20	.87	.86
alfa coëff. (subtests)	.80	.80

De onderlinge correlatie tussen de scores op vorm A en B was .84. Daar de MCCT-vormen bij dezelfde leerlingen waren afgenomen als de FRCT-vormen kon de correla-

tiecoëfficiënt worden berekend tussen deze vormen. De correlatie van de beide MCCT-vormen met de beide FRCT-vormen varieerde bij een exacte woordscoring van de FRCT-vormen tussen .62 en .65 en bij een semantische scoring tussen .64 en .66. Voor de leerlingen werd de Multiple Choice Cloze Test aangeduid als de zoektest.

4.5. Ontwikkeling van de programma's

Zoals in hoofdstuk III is vermeld, wordt bij de opzet of het ontwerp van de drie programma's zo veel mogelijk gebruik gemaakt van de resultaten van de literatuurstudie die in verband met de specificatie van de drie leestaken is uitgevoerd en van gegevens van onderzoek en aanwijzingen van ontwikkelingswerk op het gebied van de betreffende taken. Verder wordt in grote lijnen de procedure gevolgd die Gagné (1974; 1977 b) heeft voorgesteld om de leerstructuur van een te leren taak vast te stellen. Alleen de belangrijkste deeltaken of doelstellingen die de leerlingen achtereenvolgens moeten leren beheersen of bereiken, worden mede aan de hand van deze procedure geïdentificeerd.

4.5.1. Ontwikkeling van het programma Hoofdgedachte

Zoals reeds eerder vermeld, was het de bedoeling een programma van ongeveer 10 onderwijsuren te ontwikkelen dat gericht was op het leren afleiden van de hoofdgedachte uit een informatieve tekst. In het volgende zullen we een beschrijving geven van het doel en de opzet van het programma, de ontwikkeling van de eerste en tweede versie en de structuur van de definitieve versie. Voor meer informatie verwijzen we naar Deelrapport 4 (Aarnoutse et al, 1979).

Doel en opzet van het programma

Het doel van het programma wordt vrij nauwkeurig omschreven in de doelstelling van de laatste les van het programma (definitieve versie). Deze doelstelling luidt als volgt: "De leerlingen kunnen zelfstandig de hoofdgedachte afleiden uit teksten, die handelen over één thema. Bovendien kunnen de leerlingen de niet belangrijke zinnen in deze teksten opsporen. De teksten hebben een omvang van 4 tot 8 zinnen". Met de term "zelfstandig afleiden"

wordt in dit verband bedoeld, dat de leerlingen zonder hulp de hoofdgedachte kunnen afleiden en in eigen woorden kunnen formuleren. Met het "thema" van een tekst wordt hier het onderwerp bedoeld, terwijl met de term "niet belangrijke zinnen" die zinnen worden aangeduid die geen relevante informatie bevatten met betrekking tot de hoofdgedachte en de specifieke uitspraken in de tekst. Opgemerkt moet worden dat het doel van het programma verder reikt dan de taak die de leerlingen bij de Hoofdgedachtetest moeten uitvoeren. Bij deze test behoeven de leerlingen de hoofdgedachte niet zelf te formuleren en wordt hen ook niet gevraagd de niet belangrijke zinnen in een tekst aan te wijzen.

Voor de opzet van het programma werd een literatuurstudie verricht en een analyse uitgevoerd van enkele Nederlandse en buitenlandse leesprogramma's. Uit de literatuurstudie en de analyse van leesprogramma's werd duidelijk dat er allerlei opvattingen bestaan over het begrip hoofdgedachte en over procedures voor het leren bepalen van de hoofdgedachte van een tekst. Over enkele in dit verband belangrijke kwesties bleek echter een vrij grote consensus te bestaan. Zo bestaat er weinig verschil van mening over de vraag of het bepalen van het thema fundamenteel is voor het vaststellen van de hoofdgedachte van een tekst. Ook gaan de meeste auteurs van de veronderstelling uit dat het bepalen van het thema gemakkelijker is dan het bepalen van de hoofdgedachte. Tenslotte bestaat er weinig verschil van mening ten aanzien van de stelling dat het bepalen van de hoofdgedachte gemakkelijker is als de hoofdgedachte (zin) in de tekst voorkomt, dan wanneer dat niet het geval is. De drie bovengenoemde ideeën zijn bepalend geweest voor de opzet van het programma, temeer daar ze voor een deel door empirisch onderzoek worden ondersteund (vgl. Otto & Barrett, 1968; Otto, Barrett & Koenke, 1968; Van Blaricom & White, 1975, 1976). Op grond hiervan kwam de volgende opzet-in-hoofddlijnen tot stand:

- A. Het leren identificeren van het thema van een tekst.
- B. Het leren identificeren van de hoofdgedachte van een tekst.
- C. Het leren afleiden van de hoofdgedachte van een tekst.

De term "identificeren" betekent in dit verband dat het thema en de hoofdgedachte expliciet in de tekst worden vermeld en door de leerlingen worden herkend. De term "afleiden" impliceert dat de hoofdgedachte niet expliciet in de tekst voorkomt en door de leerlingen moet worden vastgesteld en ge-

formuleerd.

Gedeeltelijk in navolging van het Wisconsin Design for Reading Skill Development werd besloten om bij elk van de drie genoemde onderdelen (A, B en C) van het programma te beginnen met teksten waarvan alle zinnen relevant zijn om vervolgens ook teksten aan te bieden met enkele niet belangrijke zinnen. Hierbij werd uitgegaan van de veronderstelling dat het bepalen van het thema en de hoofdgedachte van een tekst gemakkelijker is, wanneer alle zinnen relevant zijn, dan wanneer ook enkele niet belangrijke zinnen in de tekst voorkomen. Het idee om ook teksten op te nemen waarvan niet alle zinnen relevant zijn, komt voort uit het feit dat kinderen (en volwassenen) vaak worden geconfronteerd met teksten, die niet volgens "de regels" zijn geschreven.

In afwijking van het Wisconsin Design for Reading Skill Development werd besloten bij het leren identificeren en afleiden van de hoofdgedachte de leerlingen geen teksten aan te bieden met meerkeuze vragen. Bij het leren afleiden van de hoofdgedachte staat het genereren en formuleren van de hoofdgedachte centraal. Wat betreft de omvang van de teksten werd tenslotte besloten om bij elk onderdeel korte teksten aan te bieden, variërend van twee tot acht zinnen. Hiermee was de opzet van het programma in grote lijnen gereed.

Ontwikkeling van versie 1 en 2

De opzet van het programma werd vervolgens geconcretiseerd in een reeks van tien door de leerlingen uit te voeren activiteiten en na bijstelling op enkele punten omgezet in de eerste versie van het programma. Deze versie bevatte in totaal tien lessen. Voor elke les werd een uitvoerige handleiding geschreven voor de instructeur, terwijl voor de leerlingen afzonderlijk materiaal in de vorm van teksten en opdrachten werd samengesteld. De handleiding voor elke les bevatte een beschrijving van de doelstelling van de les, de opzet van de les, het benodigde materiaal, en aanwijzingen voor de te ontwikkelen leer- en onderwijsactiviteiten. Deze aanwijzingen, die samen met de daarbij behorende teksten het grootste gedeelte van de handleiding vormden, zijn nogal gedetailleerd. Dit was nodig in verband met de functie die het programma in het onderzoek zou moeten vervullen nl. een precieze leidraad vormen voor de verschillende instructeurs. Voor een be-

schrijving van de eerste versie van het programma wordt verwezen naar Deel-rapport 4 (Aarnoutse et al., 1979).

De eerste versie van het programma werd in een vierde klas van een lage-re school in Nijmegen beproefd. Het programma werd voorafgegaan en gevolgd door respectievelijk vorm A en vorm B van de Hoofdgedachtetest. De lessen werden met een frequentie van twee per week gegeven door een student-assis-tente, die ook had meegewerkt aan de constructie van het programma. Alle lessen werden geobserveerd door de klasse-onderwijzer en een medewerker van het project. De gegevens van de voor- en nameting worden in tabel 4.7. ver-meld.

Tabel 4.7. Gegevens van de voor- en nameting met vorm A en B van de Hoofd-gedachtetest (versie 1).

	Voormeting Vorm A (okt. 1978)	Nameting Vorm B (dec. 1978)
aantal lln.	27	27
aantal items	29	29
gemid. score	11.9	18.4
stand. dev.	6.6	6.1
gemid. p.	.41	.63
KR-20	.89	.86

De gegevens vermeld in tabel 4.7. moeten met de nodige voorzichtigheid worden geïnterpreteerd, omdat deze betrekking hebben op slechts één proef-klas van 27 leerlingen, terwijl een controlegroep ontbreekt. Hoewel de prestaties van de leerlingen van de betreffende proefklas bij de nameting niet zeer goed waren, waren ze ook niet teleurstellend.

Tijdens de observatie van de lessen werden er geen indicaties gevonden die een fundamentele wijziging van de opzet van het programma noodzakelijk maakten. Dit betekende overigens niet dat het programma niet voor verbete-ring vatbaar was. Bij bijna alle lessen werden kanttekeningen en opmerkin-gen geplaatst.

Bij de constructie van de tweede versie van het programma werd zo veel

mogelijk rekening gehouden met de kanttekeningen en opmerkingen, die bij de eerste try-out werden gemaakt. Voor een beschrijving van de tweede versie van het programma wordt verwezen naar Deelrapport 4 (Aarnoutse et al., 1979).

De tweede versie van het programma werd twee weken na het begin van de eerste versie in een andere vierde klas van dezelfde lagere school in Nijmegen gegeven. Het programma werd ook nu voorafgegaan en gevolgd door respectievelijk vorm A en vorm B van de Hoofdgedachtetest. De lessen werden met een frequentie van twee per week gegeven door de onderwijzeres van de betreffende klas. Alle lessen werden geobserveerd door de student-assistente die ook meewerkte aan de constructie van het programma en een medewerker van het project.

De gegevens van de voor- en nameting worden in tabel 4.8. vermeld.

Tabel 4.8. Gegevens van de voor- en nameting met vorm A en B van de Hoofdgedachtetest (versie 2).

	Voormeting Vorm A (nov. 1978)	Nameting Vorm B (dec. 1978)
aantal lln.	25	25
aantal items	29	29
gemid. score	9.6	17.8
stand. dev.	4.4	5.1
gemid. p.	.33	.62
KR-20	.75	.80

De gegevens vermeld in tabel 4.8. moeten ook nu met de nodige voorzichtigheid worden geïnterpreteerd. Zoals uit de tabel blijkt, wijzen de gegevens in de richting van een mogelijk effect van het programma.

Uit de observatie van de lessen bleek dat het programma op verschillende punten nog voor verbetering vatbaar was. Bij de constructie van de definitieve versie werden alle opmerkingen verwerkt, die bij de tweede try-out waren gemaakt.

Tabel 4.9. Definitieve versie van het programma Hoofddgedachte.

<u>Lessen</u>	<u>Tekstkenmerken</u>	<u>Taak</u>
1. _____	relevante zinnen; thema _____ in elke zin (3-4)	thema identificeren
2. _____	relevante zinnen; thema _____ niet in elke zin (4-6)	thema identificeren
3. _____	relevante zinnen; thema _____ niet in elke zin en ver- spreid (4-9)	thema identificeren
4. _____	relevante en niet rele- vante zinnen (5-7)	thema identificeren niet relevante zin- nen opsporen
5. _____	relevante zinnen (4)	hoofddgedachte aflei- den (met instructeur)
6. _____	relevante zinnen; hoofddgedachte (zin) in de tekst (5-8)	hoofddgedachte (zin) identificeren
7. _____	relevante zinnen; hoofddgedachte (zin) in de tekst (5-8)	hoofddgedachte (zin) identificeren
8. _____	relevante en niet rele- vante zinnen; hoofdge- dachte (zin) boven de tekst (4-6)	niet relevante zin- nen opsporen
9. _____	relevante zinnen (4-9)	hoofddgedachte aflei- den
10. _____	relevante en niet rele- vante zinnen (4-8)	hoofddgedachte aflei- den; niet relevante zinnen opsporen

* De cijfers tussen haakjes geven het aantal zinnen weer.

Definitieve versie van het programma

De definitieve versie omvat 10 lessen van elk 45 minuten. In de lessen 1 tot en met 4 staat het begrip thema centraal. In deze lessen krijgen de leerlingen een strategie aangeboden voor het bepalen van het thema van korte teksten. In de lessen 5 tot en met 10 staat het begrip hoofdgedachte centraal. In deze lessen wordt de leerlingen geleerd hoe de hoofdgedachte van korte teksten kan worden vastgesteld.

Het zou te ver voeren een uitvoerige beschrijving te geven van het hele programma. We volstaan hier met een overzicht van de doelstellingen of taken die de leerlingen achtereenvolgens moeten bereiken respectievelijk beheersen. In tabel 4.9. worden deze doelstellingen of taken in het kort weergegeven. De doelstelling van de eerste les luidt (voluit): De leerlingen kunnen het thema bepalen van teksten van 3 à 4 zinnen, waarbij in elke zin het thema staat vermeld of naar het thema verwezen wordt.

Zoals eerder is opgemerkt, bestaat elke les uit een lesplan of handleiding voor de instructeur en materiaal voor de leerlingen. Elk lesplan is als volgt ingedeeld: A. Doelstelling, B. Opzet, C. Materiaal, D. Bordschema en E. Instructie. Onder E worden de concrete aanwijzingen gegeven voor de leer- en onderwijsactiviteiten die ten aanzien van de teksten kunnen worden uitgevoerd. Verder vertoont elke les ongeveer eenzelfde stramien nl. introductie of herhaling van het voorgaande, aanleren van een bepaalde strategie, inoefening van een bepaalde strategie en toepassing van de betreffende strategie. Onder de term "strategie" wordt in dit verband een manier verstaan waarop een bepaalde deeltaak kan worden uitgevoerd.

4.5.2. Ontwikkeling van het programma Bedoeling

Zoals eerder vermeld, was het de bedoeling een programma van ongeveer 10 onderwijsuren te ontwikkelen dat gericht was op het leren afleiden van de bedoeling van de hoofdpersoon uit een fictieve tekst. In het volgende wordt een beschrijving gegeven van het doel en de opzet van het programma, de ontwikkeling van de eerste en tweede versie en de structuur van de definitieve versie. Voor meer informatie verwijzen we naar Deelrapport 4 (Aarnoutse et al., 1979).

Doel en opzet van het programma

Het doel van het programma kan worden geformuleerd in de termen van de doelstelling van de laatste les van het programma (definitieve versie). Deze doelstelling luidt als volgt: "De leerling kan bij een tekst de vraag waarom een bepaalde persoon iets doet of zegt, beantwoorden door de bedoeling die deze persoon heeft te formuleren". Het gebruikte teksttype is het verhaal, waarin tenminste één persoon voorkomt die een aantal handelingen verricht als onderdeel van een plan en als reactie op een beginsituatie of -gebeurtenis. De vraag naar de bedoeling of het motief van de hoofdpersoon is een vraag naar de uiteindelijke bedoeling, voor zover die uit de tekst is af te leiden. De bedoeling van de hoofdpersoon staat niet expliciet in de tekst, maar kan door het combineren van een aantal gegevens worden afgeleid. Opgemerkt moet worden dat het doel van het programma verder reikt dan de taak die de leerlingen bij de Bedoelingtest moeten uitvoeren. Bij deze test behoeven de leerlingen de bedoeling van de hoofdpersoon niet zelf te formuleren. In het programma wordt deze taak overigens vereenvoudigd door het antwoord vóór te structureren; de leerlingen kunnen volstaan met het aanvullen van het antwoord.

Voor de opzet van het programma kon alleen gebruik worden gemaakt van studies waarin begrippen als "bedoeling" (Miller & Johnson-Laird, 1976) en "plannen" (Schank & Abelson, 1977) worden geanalyseerd en waarin de structuur van verhalende teksten wordt onderzocht (Rumelhart, 1977b). Onderzoeksgegevens en curriculummateriaal met betrekking tot het identificeren en afleiden van de bedoeling van de hoofdpersoon waren - voor zover bekend - niet voorhanden.

Bij de opzet van het programma werd in eerste instantie uitgegaan van de veronderstelling dat kennis omtrent de structuur van verhalende teksten van belang is bij het vaststellen van de bedoeling van iemand. Verder werd er van uitgegaan dat kinderen in teksten onderscheid moeten leren maken tussen de bedoeling van een persoon en de middelen en wegen die hij gebruikt om zijn doel (intentie) te realiseren. Wat betreft het vaststellen van iemands bedoeling werd tenslotte onderscheid gemaakt tussen het identificeren en afleiden van de bedoeling. Afhankelijk van de mate waarin aanwijzingen van iemands bedoeling in de tekst voorkomen, werden drie niveaus onderscheiden nl.

- a. de bedoeling staat expliciet in de tekst vermeld;
- b. de tekst bevat expliciete aanwijzingen van iemands bedoeling zoals het gebruik van verba als willen, van plan zijn en dergelijke;
- c. de tekst bevat geen expliciete (linguïstische) aanwijzingen; er worden slechts handelingen beschreven waaraan een motief moet worden toegekend.

Op het eerste niveau is sprake van het identificeren van de bedoeling, terwijl op het tweede en derde niveau gesproken kan worden van het afleiden van de bedoeling.

Op basis van de voorgaande overwegingen kwam voor de eerste versie de volgende opzet-in-hoofdlijnen tot stand:

- A. De structuur van een verhaal leren.
- B. De relaties tussen doel en middelen leren.
 - het leren identificeren van een bedoeling;
 - het leren identificeren van een plan en de uitvoering daarvan als middel om een bedoeling te realiseren.
- C. Het leren afleiden van een bedoeling.

Besloten werd bij het leren identificeren en afleiden van de bedoeling de leerlingen geen teksten aan te bieden met meerkeuze vragen. Bovendien werd besloten om de teksten geleidelijk in omvang te laten toenemen.

Ontwikkeling van versie 1 en 2

De opzet van het programma werd geconcretiseerd in een reeks van 10 lessen. Voor elke les werd een uitvoerige handleiding geschreven, die een beschrijving bevat van de doelstelling en opzet van de les, alsmede aanwijzingen voor de instructeur. Voor de leerlingen werd afzonderlijk materiaal samengesteld, dat uit teksten en te maken oefeningen bestond. Voor een beschrijving van de inhoud van de eerste versie verwijzen we naar Deelrapport 4 (Aarnoutse et al., 1979). Opgemerkt moet worden dat in deze versie een sterk accent ligt op de verhaalstructuur; veel oefeningen zijn er op gericht de leerlingen hiermee vertrouwd te maken.

De eerste versie van het programma werd in een vierde klas beproefd, waarvan de leerlingen in twee groepen waren verdeeld. De lessen werden voorafgegaan en gevolgd door respectievelijk de tweede en derde proefversie van de Bedoelingtest. De lessen werden met een frequentie van twee per week gegeven door een medewerker van het project en door een student-assistente,

die ook had meegewerkt aan de constructie van het programma. De klasleerkracht woonde steeds een van de lessen bij en gaf na afloop haar oordeel over zowel de geplande als de uitgevoerde les. De belangrijkste ervaringen werden na elke les op schrift gesteld om in een volgende versie te worden verwerkt. De gegevens van de voor- en nameting worden in tabel 4.10. vermeld.

Tabel 4.10. Gegevens van de voor- en nameting met een proefversie van de Bedoelingtest (versie 1).

	Voormeting	Nameting
aantal lln.	27	27
aantal items	37	37
gemid. score	22.2	26.9
stand. dev.	9.2	7.7
gem. p.	.60	.74

Omdat de ontwikkeling van de Bedoelingtest nog niet was afgerond, konden bij de analyse van de data alleen de items worden betrokken die niet tussentijds waren gewijzigd. Verdere reserve ten aanzien van de interpretatie van de gegevens is nodig, omdat deze slechts betrekking hebben op één proefklas van 27 leerlingen, terwijl een controlegroep ontbreekt. Met de nodige voorzichtigheid kunnen de gegevens worden geïnterpreteerd als een aanwijzing dat het programma effect heeft.

Tijdens de uitvoering van het programma werd duidelijk dat een aantal onderdelen niet aan de verwachtingen voldeed. Het beantwoorden van vragen met het doel de verhaalstructuur in schema te brengen werd bijvoorbeeld door de leerlingen niet als een hulpmiddel ervaren bij het identificeren en afleiden van de bedoeling. Bovendien bleek dat in het programma te snel werd begonnen met de taak de bedoeling af te leiden, terwijl begrippen als bedoeling of manieren om het doel te bereiken voor de leerlingen nog niet vertrouwd genoeg waren.

Bij de constructie van de tweede versie van het programma werd zo veel

mogelijk rekening gehouden met de ervaringen opgedaan tijdens de eerste try-out. Geprobeerd werd de relevante begrippen en de daarmee corresponderende typen van vragen geleidelijker in te voeren. Besloten werd om slechts één nieuw begrip (of een nieuw vraagtype) per les te introduceren.

De tweede versie van het programma werd gegeven in één vierde en één gecombineerde vierde en vijfde klas. Het programma werd voorafgegaan en gevolgd door respectievelijk vorm A en vorm B van de Bedoelingtest. De lessen werden gegeven met een frequentie van twee per week. In een klas werd het programma aangeboden door de klasse-onderwijzeres, in de andere klas door een student-assistente. De lessen werden bijgewoond door een medewerker van het project en een student-assistente, die ook meewerkte aan de ontwikkeling van het programma. Bovendien woonde de klasse-onderwijzeres de lessen bij, die door de student-assistente werden gegeven. De belangrijkste ervaringen werden na elke les op schrift gesteld. De gegevens van de voor- en nameting worden in tabel 4.11. vermeld.

Tabel 4.11. Gegevens van de voor- en nameting met vorm A en B van de Bedoelingtest (versie 2).

	Voormeting Vorm A	Nameting Vorm B
aantal lln.	60	60
aantal items	24	24
gemid. score	15.2	17.4
stand. dev.	4.5	4.2
gem. p.	.63	.73

De gegevens vermeld in tabel 4.11. hebben betrekking op alle leerlingen. Het programma is echter bestemd voor leerlingen van het vierde leerjaar. Wanneer we alleen de scores van deze leerlingen nemen, krijgen we een iets ander resultaat (zie tabel 4.12.).

Tabel 4.12. Gegevens van de voor- en nameting met Vorm A en B van de Bedoelingtest (versie 2, leerlingen van het vierde leerjaar).

	Voormeting Vorm A	Nameting Vorm B
aantal lln.	42	42
aantal items	24	24
gemid. score	14.4	16.8
stand. dev.	4.1	4.4
gem. p.	.60	.70

Met de nodige voorzichtigheid kunnen deze gegevens worden geïnterpreteerd als een aanwijzing dat het programma mogelijk effect heeft gehad, zij het dat dit effect niet groot is. Bij de uitvoering bleek dat het programma op een aantal punten nog verbeterd kon worden.

Definitieve versie van het programma

Bij de constructie van de definitieve versie werd geprobeerd de tekorten van de tweede versie op te heffen. De veranderingen hadden vooral betrekking op de vragen en opdrachten die meer op de eindtaak moesten worden afgestemd en op de structuur van de lessen. In de lessen werden systematisch vragen geïntroduceerd, die het begrip bedoeling trachten te verduidelijken zoals:

- a. wat is de bedoeling van iemand?
- b. wat doet iemand om het doel te bereiken?
- c. wordt het doel bereikt?
- d. waarom heeft iemand een bedoeling?

Wat betreft de structuur kregen alle lessen ongeveer een gelijke opzet nl. een herhaling van de inhoud van de vorige les en een zelfstandig te maken oefening, een introductie van het probleem aan de hand van gezamenlijk uit te voeren opdrachten en twee zelfstandig te maken opdrachten.

De definitieve versie omvat 10 lessen van elk 45 minuten. In de lessen 1 tot en met 5 staan de vier vragen die het begrip bedoeling karakteriseren centraal. De lessen 6 tot en met 10 vormen een verdere uitwerking van

deze vragen. Het zou te ver voeren om hier een uitvoerige beschrijving te geven van het hele programma. We volstaan hier met een beschrijving van het doel van elke les (zie tabel 4.13).

Tabel 4.13. Definitieve versie van het programma Bedoeling.

- Les 1. De leerling kan bij een eenvoudige tekst, waarin de bedoeling van iemand expliciet wordt vermeld, antwoord geven op de vraag wat iemand wil of beoogt.
- Les 2. De leerling kan bij een eenvoudige tekst, waarin onder andere de handelingen van iemand worden beschreven, antwoord geven op de vraag wat iemand doet om zijn doel te bereiken.
- Les 3. De leerling kan bij een eenvoudige tekst, waarin onder andere de afloop van een reeks handelingen wordt beschreven, antwoord geven op de vraag of iemand zijn doel bereikt of niet.
- Les 4. De leerling kan bij een eenvoudige tekst waarin onder andere een gebeurtenis wordt beschreven die de aanleiding vormt voor een bepaald motief, antwoord geven op de vraag hoe het komt dat iemand een bepaalde bedoeling heeft.
- Les 5. De leerling kan bij een tekst, waarin onder andere een aantal tussendoelen worden beschreven, antwoord geven op de vraag naar de uiteindelijke bedoeling.
- Les 6. De leerling kan bij een eenvoudige tekst de bedoeling van de hoofdpersoon afleiden. De bedoeling wordt niet expliciet in de tekst vermeld.
- Les 7. De leerling kan bij een tekst, waarin een beschrijving wordt gegeven van een plan en de uitvoering daarvan, de bedoeling van de hoofdpersoon afleiden. De bedoeling staat niet expliciet in de tekst vermeld.
- Les 8. De leerling kan bij een tekst, waarin een list wordt beschreven, de bedoeling van de hoofdpersoon afleiden. De bedoeling staat niet expliciet in de tekst.
- Les 9. De leerling kan bij een tekst, waarin twee personen worden beschreven die tegenstrijdige belangen hebben, de bedoeling van de hoofdpersonen afleiden.

Les 10. De leerling kan bij een tekst de vraag waarom een bepaalde persoon iets doet of zegt, beantwoorden door de bedoeling die deze persoon heeft, te formuleren.

4.5.3. Ontwikkeling van het programma Redeneren

In het kader van het onderzoek werd een programma van ongeveer 10 onderwijssuren ontwikkeld dat gericht is op het leren uitvoeren van de eerder omschreven redeneertaak. In het volgende zullen we een beschrijving geven van het doel en de opzet van het programma, de ontwikkeling van de eerste en tweede versie en de structuur van de definitieve versie. Voor meer informatie verwijzen we naar Deelrapport 4 (Aarnoutse et al., 1979).

Doel en opzet van het programma

Het programma heeft tot doel de leerlingen van het vierde leerjaar te leren een conclusie-vraag te beantwoorden over een tekst, die een major en een minor bevat. Het doel van het programma wordt vrij nauwkeurig omschreven in de doelstelling van de laatste les van het programma (definitieve versie). Deze doelstelling luidt als volgt: "De leerlingen kunnen uit een aantal zinnen in tekstvorm de twee zinnen (i.c. de major en de minor) selecteren die nodig zijn voor het antwoord op de conclusie-vraag en het juiste antwoord op de conclusie-vraag kiezen uit drie voorgegeven antwoorden". Opgemerkt moet worden dat het doel van het programma niet verder reikt dan de taak die de leerlingen bij de Redeneertest moeten uitvoeren.

Voor de opzet van het programma Redeneren werd gebruik gemaakt van de onderzoeksgegevens van onder anderen Ennis en Paulus (1965), Roberge (1970), Carroll (1975) en Ennis (1976) en in beperkte mate van het redeneerprogramma van het Wisconsin Design.

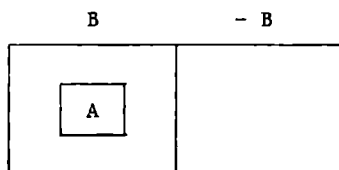
Bij de opzet van het programma werden twee deeltaken aan de redeneertaak onderscheiden nl. een selectie-taak en een logica-taak. Bij de uitvoering van de taak moeten de leerlingen eerst de relevante zinnen uit een tekst selecteren. Deze zinnen zijn de major, de minor en in sommige gevallen de $B > A$ zin. Vervolgens moeten ze met de geselecteerde zinnen logisch juist redeneren om de conclusie-vraag te kunnen beantwoorden.

Het belangrijkste uitgangspunt bij de opzet van het programma was het

streven de logica-taak, de redenering, te visualiseren. Zoals reeds eerder is opgemerkt, bestaat een major uit een A-term en een B-term. De termen van zowel de categorische als van de hypothetische major kunnen worden opgevat als verwijzend naar verzamelingen. Omdat verzamelingen concreet aanwezig kunnen zijn of afgebeeld kunnen worden, liggen hier de gewenste visualiseringsmogelijkheden.

Een eerste vraag was nu wat van een redenering moet worden afgebeeld. Een beschrijving van de delen van een redenering in termen van verzamelingen geeft hiervoor aanwijzingen. Een categorische of hypothetische major stelt dat A een deelverzameling is van B. Zowel in de test als in het programma is het steeds duidelijk dat B groter is dan A. Een minor geeft aan dat een bepaald element lid is van de verzameling A, - A, B of - B. Een conclusie-vraag vraagt of dit bepaalde element lid is van de verzameling A of B. Deze beschrijving suggereert dat een tekening in ieder geval de verzamelingen A, - A, B en - B in hun onderlinge relaties moet bevatten. Hoe die onderlinge relaties zijn, wordt aangegeven door de major. Verder moet in de tekening duidelijk kunnen worden aangegeven in welk gebied het minor element zich bevindt. Tenslotte moet het antwoord op de conclusie-vraag uit de tekening kunnen worden afgelezen.

Een volgende vraag was hoe een tekening er uit moet zien, rekening houdend met de bovengenoemde eisen. Hiervoor zijn verschillende mogelijkheden. Voor de eerste versie van het programma werd gekozen voor de onderstaande tekening.



Een derde vraag was welke stappen bij het maken van een tekening van een redenering onderscheiden moeten worden. Een tekening kan gemaakt worden op basis van een major alleen. (Een categorische en een hypothetische major leiden tot een zelfde tekening). In de major worden echter niet alle verzamelingen expliciet genoemd, die in de tekening wel moeten worden geëxplici-

teerd (bijvoorbeeld de verzamelingen - A en - B en de wederzijdse uitsluiting van de verzamelingen A en - B). Nu is het mogelijk teksten te construeren die verzamelingen en relaties daarbinnen, die in de major impliciet blijven, expliciteren. Dergelijke teksten met uitgeschreven verzamelingen bevatten óók een major. In de eerste versie van het programma werd van deze teksten gebruik gemaakt. Hierbij werd verondersteld dat het maken van een tekening op basis van een tekst met uitgeschreven verzamelingen faciliterend zou werken op het maken van een tekening op basis van een major alleen. Wat betreft de aard van de tekening werd bij de opbouw van het programma onderscheid gemaakt tussen een ongeordende tekening met elementen, een geordende tekening met elementen en een geordende tekening zonder elementen. Verondersteld werd dat het lezen van een geordende tekening zonder elementen vergemakkelijkt kan worden wanneer de twee andere soorten tekeningen daaraan vooraf gaan.

Zoals reeds eerder vermeld, bestaat de redeneertaak niet alleen uit een logica-taak maar ook uit een selectie-taak. Wat betreft de selectie-taak werden de volgende twee stappen onderscheiden:

- a. De tekst bevat alleen een major. De minor wordt apart gegeven samen met de conclusie-vraag. De taak bestaat uit het selecteren van de major in de tekst;
- b. de tekst bevat een major en een minor. De taak is beide te selecteren, gegeven een conclusie-vraag.

Zoals uit het voorgaande blijkt, werd eerst een ontwerp gemaakt voor de twee onderscheiden deeltaken. Deze twee ontwerpen werden vervolgens in elkaar geschoven tot één opzet-in-hoofdlijnen.

Ontwikkeling van versie 1 en 2

De opzet van het programma werd geconcretiseerd in een reeks van 10 lessen. Voor elke les werd een uitvoerige handleiding geschreven, die een beschrijving bevat van de doelstelling en opzet van elke les, alsmede aanwijzingen voor de instructeur. Voor de leerlingen werd afzonderlijk materiaal samengesteld, dat uit tekeningen, teksten en opdrachten bestond. Voor een beschrijving van de inhoud van de eerste versie verwijzen we naar Deelrapport 4 (Aarnoutse et al., 1979).

De eerste versie van het programma werd beproefd in één vierde klas van

een lagere school in Boxtel. De lessen werden gegeven door een student-assistent die ook had meegewerkt aan de ontwikkeling van het programma. Alle lessen werden geobserveerd door en na afloop besproken met de klasse-onderwijzer. Enkele lessen werden ook geobserveerd door een medewerker van het project. Na afloop van het programma werd een deel van de Redeneertest (deel A1) afgenomen. De gegevens die deze test opleverde, moeten met de nodige voorzichtigheid worden geïnterpreteerd, omdat ze betrekking hebben op slechts één klas leerlingen en verzameld zijn met de helft van een test. Bovendien ontbreekt een voormeting en een controlegroep. De prestaties van de leerlingen van de proefklas werden vergeleken met de scores van ± 160 andere leerlingen bij de afname van dezelfde 20 items ten behoeve van de constructie van de parallelvormen. Hoewel de meting in de proefklas ruim een half jaar vroeger in het schooljaar plaats vond nl. in november in plaats van in juni, was het gemiddelde percentage goed van deze klas hoger, zij het dat het verschil niet groot was.

Tabel 4.14. Percentage goed beantwoorde items van deel A1 van de Redeneertest van de proefklas in Boxtel en een vergelijkingsgroep (versie 1).

	Proefklas Boxtel november 1978	Constructie parallelvormen juni 1978
Totaal (20 items)	74%	65%
Subtest Zeker (12 items)	74%	73%
Subtest Onzeker (8 items)	75%	54%

De gegevens van tabel 4.14. maken het waarschijnlijk, dat het programma effect heeft en dat het effect het sterkst is op de subtest "Onzeker".

De gang van zaken in de proefklas was minder bevredigend dan wellicht uit de resultaten van de nameting verwacht zou worden. Tijdens de try-out werd duidelijk dat een aantal onderdelen niet aan de verwachtingen voldeed. De leerlingen vonden het bijvoorbeeld moeilijk om de tekening te lezen.

Verder bleek de selectie van de relevante zinnen uit een tekst een belangrijk probleem te zijn voor de leerlingen. Bovendien bleek dat de teksten met uitgeschreven verzamelingen eerder verwarrend dan verhelderend werkten.

Bij de constructie van de tweede versie van het programma werd een aantal ingrijpende wijzingen in het programma aangebracht. De tekening werd gewijzigd in een 2 x 2-matrix, terwijl aan de selectie van de relevante zinnen meer en volgens een andere systematiek aandacht werd besteed (zie beschrijving van de definitieve versie van het programma). Bovendien werden de teksten met uitgeschreven verzamelingen uit het programma geschrapt.

De tweede versie van het programma werd gegeven in één vierde klas. Het programma werd voorafgegaan en gevolgd door respectievelijk vorm A en B van de Redeneertest. De leerlingen werden in twee groepen verdeeld. De ene groep kreeg les van de student-assistent die ook meewerkte aan de ontwikkeling van het programma. De andere groep, die enkele dagen later aan het programma begon, kreeg les van de klasse-onderwijzer. Deze laatste serie lessen werd bijgewoond door een medewerker van het project. Elke les van de klasse-onderwijzer werd direkt na afloop besproken. De belangrijkste ervaringen werden ook nu genoteerd.

Tabel 4.15. toont aan dat de leerlingen zeer sterk vooruit zijn gegaan. Hoewel het aantal leerlingen klein is en een controlegroep ontbreekt, lijkt het programma duidelijk effect te hebben. Overigens vertoont ook deze proefklas, evenals de proefklas bij de eerste versie, de grootste vooruitgang op de subtest Onzeker.

Tabel 4.15. Percentage goed beantwoorde items bij de voor- en nameting met vorm A en B van de Redeneertest (versie 2).

	Voormeting Vorm A	Nameting Vorm B
Totaal (40 items)	54%	84%
subtest Zeker (24 items)	57%	83%
subtest Onzeker (16 items)	50%	85%

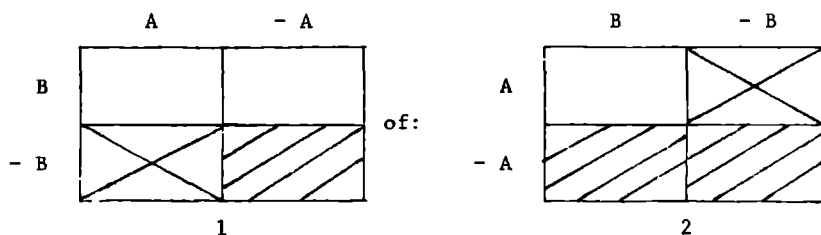
Tijdens de uitvoering van de tweede versie bleek dat het programma op een aantal niet essentiële punten nog voor verbetering vatbaar was.

Definitieve versie van het programma

Het programma bestaat uit twee gedeeltelijk parallel lopende delen nl. een tekeningdeel en een selectiedeel. In het tekeningdeel leren de kinderen een tekening maken van een major en een minor zodat ze het antwoord op de conclusie-vraag kunnen aflezen. In het selectiedeel leren de kinderen die zinnen te selecteren, die nodig zijn om de conclusie-vraag te beantwoorden nl. de major en de minor en in sommige gevallen ook de $B > A$ -zin. In het volgende geven we een korte beschrijving van beide delen.

Tekeningdeel

De tekening die de leerlingen van een redenering leren maken, ziet er als volgt uit:



De tekening is gebaseerd op de categorische of hypothetische major. Er is altijd één hok dat leeg is nl. $A, -B$. Dit hok is niet mogelijk, omdat: alle A zijn B, of: als A dan B. In dit hok wordt in de loop van het programma een kruis gezet. De drie andere hokken zijn wel mogelijk.

In de tekening wordt de minor aangegeven door één of twee hok(ken) te arceren. In tekening 1 is minor $-B$ aangegeven en in tekening 2 minor $-A$. De tekening wordt steeds ontwikkeld uitgaande van een conclusie-vraag, waarbij een major en een minor zijn gegeven. De vraag "A?" leidt tot tekening 1 en de vraag "B?" tot tekening 2. Uit de tekening kan het antwoord worden afgelezen. In tekening 1 is het antwoord: nee ($-A$) en in tekening 2 is het antwoord: onzeker (B of $-B$).

Een voorbeeld van tekening 1 is:

Alle schriften van Karel zijn groen (major)
 Dit schrift is niet groen (minor - B)
 Is dit schrift van Karel? (conclusie-vraag)

	van Karel	niet van Karel	
groen			schriften
niet groen			

Een voorbeeld van tekening 2 is:

Alle schriften van Karel zijn groen (major)
 Dit schrift is niet van Karel (minor - A)
 Is dit schrift groen? (conclusie-vraag)

	groen	niet groen	
van Karel			schriften
niet van Karel			

Het tekeningendeel is als volgt opgebouwd:

In les 1 en 2 worden oefeningen gegeven in het lezen van een 2 x 2-matix. In les 3 en 4 demonstreert de leerkracht de ontwikkeling van de tekening met elementen. De leerlingen arceren het minor-gebied en lezen het antwoord af. In les 3 komen categorische majors aan de orde en in les 4 hypothetische majors.

In les 5 en 6 zetten de leerlingen een kruis in het hok dat door de major wordt uitgesloten (A, -B). Hierbij is gegeven een categorische of hypothetische major en een tekening zonder elementen maar met namen.

In les 7 en 8 schrijven de leerlingen namen bij een tekening, zetten een kruis in het hok dat door de major wordt uitgesloten, arcieren het minor-gebied en lezen het antwoord af. Hierbij is gegeven een major, een minor, een conclusie-vraag en een reeks stapsgewijze opdrachten.

In les 9 en 10 wordt de reeks stapsgewijze opdrachten verkort en moeten de leerlingen eerst de major en de minor selecteren.

Selectiedeel

Om een conclusie-vraag te kunnen beantwoorden, hebben de leerlingen de major en de minor nodig en in sommige gevallen ook de $B > A$ -zin. Het programma bevat geen algoritme voor de selectie van zinnen die nodig zijn voor het antwoord op de conclusie-vraag. Uitgangspunt is dat de leerlingen door vaak selectie-taken uit te voeren bij teksten die geleidelijk langer worden, een zekere feeling ontwikkelen voor de zinnen die nodig zijn voor het antwoord.

De selectie-taak houdt in dat steeds de twee zinnen die nodig zijn voor het antwoord op de conclusie-vraag onderstreept worden. De teksten waaruit de twee zinnen geselecteerd moeten worden, vertonen een bepaalde opbouw. Het aantal irrelevante zinnen neemt geleidelijk aan toe.

4.6. Selectie van de scholen en vergelijking van de experimentele en controlescholen

In deze paragraaf wordt een beschrijving gegeven van de procedure die gevolgd is bij de selectie van de scholen. Verder wordt aandacht besteed aan de vraag in hoeverre de experimentele en controlescholen met elkaar vergelijkbaar zijn wat betreft de vormgeving van het leesonderwijs.

4.6.1. Selectie van de scholen

Voor het onderzoek waren in totaal acht grote lagere scholen nodig; vier scholen waar de programma's gegeven zouden worden (experimentele scholen) en vier scholen waar geen programma's zouden worden gegeven (controlescholen). De scholen die voor selectie in aanmerking kwamen, moesten aan twee voorwaarden voldoen. In de eerste plaats dienden de scholen "normale" lagere scholen te zijn. Dit betekende dat scholen met een bijzonder karakter zoals Jenaplan-scholen, stimuleringsscholen en vernieuwingsscholen niet

voor selectie in aanmerking kwamen. In de tweede plaats dienden de scholen over twee klassen voor het vierde leerjaar te beschikken en over twee extra lokalen die tijdelijk gebruikt konden worden. Verondersteld werd dat scholen met 12 of meer leerkrachten en een bestand van meer dan 300 leerlingen eerder aan de laatst genoemde voorwaarde zouden voldoen dan scholen met minder leerkrachten en minder leerlingen.

Het plan was om een gestratificeerde steekproef van scholen, met random trekking uit de strata, samen te stellen uit de verzameling grote lagere scholen uit Nijmegen en omgeving. Bij de selectie van de scholen zou een keuze-volgorde worden vastgesteld. Pas na de selectie zou met de hoofden van de betreffende scholen contact worden opgenomen en worden nagegaan of de scholen aan de bovengenoemde voorwaarden voldeden en of het hoofd en de leerkrachten van de vierde klas van deze scholen bereid waren aan het onderzoek mee te doen. In het volgende wordt een beschrijving gegeven van de procedure die bij de selectie van de scholen is gevolgd.

De eerste taak die moest worden uitgevoerd, was de bepaling van het gebied waaruit de scholen geselecteerd zouden worden. Om praktische en financiële redenen werd gekozen voor het voormalige stadsgewest Nijmegen inclusief één gemeente ten noorden van dit gewest. Tot het stadsgewest behoorden de volgende gemeenten: Nijmegen, Bommel, Bergharen, Beuningen, Druten, Ewijk, Groesbeek, Heumen, Millingen aan de Rijn, Overasselt, Ubbergen en Wijchen. De gemeente ten noorden van het gewest was Elst.

De tweede taak was het inwinnen van informatie omtrent de grootte van de scholen in de betreffende gemeenten. Alle gemeenten werden telefonisch benaderd en gevraagd informatie te verschaffen omtrent het aantal scholen voor gewoon lager onderwijs binnen de gemeente, het aantal leerlingen per school en het aantal leerkrachten. Verschillende gemeenten buiten Nijmegen konden geen exacte informatie geven omtrent het aantal leerkrachten per school. Na deze inventarisatie kon worden vastgesteld dat in de gemeente Nijmegen 9 scholen waren met 361 of meer leerlingen en 12 of meer leerkrachten. Eén school nl. een Jenaplanschool viel uit vanwege het bijzondere karakter van deze school. Wat de andere gemeenten betreft waren er in totaal 17 scholen met een leerlingenaantal van 312 of meer. Besloten werd deze 17 scholen voor selectie aan te houden en in een latere fase na te gaan of ze aan de voorwaarden voor deelname aan het onderzoek konden vol-

doen. Verondersteld werd dat scholen uit buitengemeenten eerder in staat waren aan het onderzoek te participeren dan scholen in Nijmegen, die vaak al "belast" zijn met studenten van pedagogische academies of deelnemen aan vernieuwingsactiviteiten van de schooladviesdienst.

De derde taak was de indeling van de verzameling grote lagere scholen uit Nijmegen en omgeving volgens een bepaald criterium. Besloten werd de scholen in te delen op grond van de urbanisatiegraad van de gemeenten. Overeenkomstig de C.B.S.-publicatie "Typologie van de Nederlandse gemeenten naar urbanisatiegraad, 31 mei 1960" werden drie typen van gemeenten onderscheiden nl. gemeenten met een stedelijk karakter, verstedelijkte plattelandsgemeenten en plattelandsgemeenten. Het zal duidelijk zijn dat de scholen afkomstig uit Nijmegen werden gerangschikt onder de categorie "scholen uit gemeenten met een stedelijk karakter". De scholen uit Druten, Groesbeek, Millingen aan de Rijn, Wijchen en Ubbergen werden overeenkomstig de C.B.S.-publicatie "Bevolking der gemeenten van Nederland op 1 januari 1975" geplaatst in de categorie "scholen uit verstedelijkte plattelandsgemeenten". De scholen uit Bemmelen, Bergharen, Beuningen, Ewijk, Heumen, Overasselt en Elst werden op grond van de C.B.S.-publicatie van 1975 gerangschikt onder de categorie "scholen uit plattelandsgemeenten". Van het totaal aantal grote lagere scholen (25) vielen er op grond van deze typologie van gemeenten 8 scholen in de categorie "scholen uit gemeenten met een stedelijk karakter", 10 scholen in de categorie "scholen uit verstedelijkte plattelandsgemeenten" en 7 scholen in de categorie "scholen uit plattelandsgemeenten". Besloten werd om voor het onderzoek 2 scholen uit de eerste categorie te selecteren, 4 scholen uit de tweede categorie, en 2 scholen uit de laatstgenoemde categorie. Het aantal scholen per categorie werd vrij willekeurig bepaald.

De vierde taak was de selectie van de scholen die als experimentele scholen en controlescholen in aanmerking zouden komen. Per categorie werden de scholen ad random in twee groepen verdeeld nl. een groep experimentele scholen en een groep controlescholen en met behulp van een tabel met random getallen in een volgorde geplaatst. In geval een school niet aan de voorwaarden kon voldoen of niet bereid was in het onderzoek te participeren zou de volgende school gevraagd worden.

Na de selectie van acht lagere scholen werd met de hoofden van deze scho-

len contact opgenomen. Na een mondelinge uiteenzetting over het doel en de opzet van het onderzoek waren de hoofden en de leerkrachten van de vierde klassen bereid aan het onderzoek deel te nemen. Twee scholen bleken vanwege hun bijzondere karakter niet aan de voorwaarden te voldoen. De scholen die vervolgens in aanmerking kwamen, voldeden aan de voorwaarden en waren tevens bereid aan het onderzoek mee te doen. De besturen van de betrokken scholen verleenden schriftelijk toestemming tot uitvoering van het onderzoek. De inspecteurs van het kleuteronderwijs en van het lager onderwijs in de inspectie Nijmegen, Wijchen en Elst verleenden aan de betrokken scholen toestemming tot afwijking van het rooster voor zover dat in verband met het onderzoek noodzakelijk was.

Hoewel het aanvankelijk de bedoeling was om twee vierde klassen van elke school in het onderzoek te betrekken (totaal aantal klassen: 16) werd op verzoek van het hoofd en de leerkrachten van één experimentele school en twee controlescholen nog één klas van deze scholen bij het onderzoek betrokken. Dit betekende dat in totaal 19 klassen aan het onderzoek zouden deelnemen. Wat betreft de namen van de scholen, de hoofden van scholen en de leerkrachten wordt verwezen naar Bijlage 4.1.

4.6.2. Vergelijking van de experimentele en controlescholen

Zoals reeds eerder vermeld, wordt in het onderzoek onder andere het verschil getoetst tussen de prestaties van de leerlingen van de experimentele scholen en de controlescholen op een vijftal tests die tijdens de nameting en retentiemeting worden afgenomen. Bij de interpretatie van eventuele verschillen kunnen de gemiddelde prestaties en dergelijke van de leerlingen van beide groepen scholen op de vijf tests die tijdens de voormeting worden afgenomen, betrokken worden. Een geconstateerd verschil kan met meer recht toegeschreven worden aan experimentele condities wanneer kan worden aangenomen dat er geen systematische verschillen bestaan tussen beide groepen scholen wat betreft de praktijk van het leesonderwijs. Om een beeld te krijgen van de vergelijkbaarheid van beide groepen van scholen wat betreft de vormgeving van het leesonderwijs, werd een vragenlijst ontwikkeld. Deze vragenlijst werd in juni 1979 aan de leerkrachten van de vierde klas van de experimentele scholen en de controlescholen ter invulling voorgelegd.

De vragenlijst werd ontwikkeld door een doctoraal studente in de onder-

wijskunde. De projectmedewerkers fungeerden als begeleiders. De ontwikkeling van de vragenlijst verliep als volgt. Na de bestudering van de nodige literatuur op het gebied van lezen en leesonderwijs en diverse leesmethoden en schoolwerkplannen van lagere scholen werd een eerste versie geschreven. Deze versie werd door een aantal leerkrachten van een vierde, vijfde en zesde klas ingevuld en van commentaar voorzien. Van de kant van de projectmedewerkers werd ook een aantal opmerkingen c.q. suggesties gegeven. Voor de ontwikkeling van de tweede versie werden in totaal 14 leeslessen op drie verschillende scholen bijgewoond. Van elke les werd door de betreffende studente een verslag gemaakt en naar aanleiding daarvan enkele vragen geformuleerd die mogelijk gebruikt konden worden voor de tweede versie. De tweede versie werd door vier leerkrachten van vierde klassen ingevuld. Uit deze laatste proefafname bleek dat de invulling geen problemen opleverde en dat de benodigde invultijd \pm 60 minuten bedroeg.

De definitieve versie van de vragenlijst bestaat uit de volgende onderdelen:

- a. vragen over de school, het aantal uren leesonderwijs, de gebruikte leesmethoden en de bibliotheekvoorzieningen;
- b. vragen met betrekking tot het technisch lezen;
- c. vragen met betrekking tot het begrijpend lezen;
- d. vragen met betrekking tot studerend lezen;
- e. vragen met betrekking tot belangstelling voor lezen.

Opgemerkt moet worden dat met behulp van de antwoorden op deze vragenlijst niet meer dan een algemene indruk kan worden verkregen wat betreft het leesonderwijs in beide groepen van scholen. Om een gefundeerd oordeel te kunnen geven omtrent de vergelijkbaarheid van beide groepen van scholen wat betreft het leesonderwijs zou de vragenlijst verder beproefd en geanalyseerd moeten worden en zou een onderzoek zoals Durkin (1978-1979) heeft uitgevoerd, noodzakelijk zijn geweest. Door het geringe aantal respondenten (19) is toetsing op significante verschillen in antwoorden tussen beide groepen van leerkrachten niet zinvol.

In het volgende zullen we elk onderdeel van de vragenlijst in het kort beschrijven en de conclusie vermelden die op grond van de antwoorden van de respondenten geformuleerd kan worden. Voor meer gedetailleerde informatie verwijzen we naar Deelrapport 5 (Aarnoutse et al., 1980).

Het eerste onderdeel van de vragenlijst heeft betrekking op meer algemene zaken als de grootte van de school, eventuele begeleiding van een schooladviesdienst, het aantal uren leesonderwijs op het lesrooster, de gebruikte methoden voor aanvankelijk en voortgezet lezen en het gebruik van klas- en schoolbibliotheek en/of documentatiecentrum. Uit de antwoorden van de leerkrachten bleek dat er geen systematische verschillen waren tussen de twee groepen van scholen wat betreft de bovengenoemde punten.

Het tweede onderdeel heeft betrekking op technisch lezen. Dit aspect van het lezen wordt in de vragenlijst omschreven als het technisch en expressief correct omzetten van geschreven of gedrukte woorden, woordgroepen en zinnen in gesproken taal. Bij dit onderdeel werd gevraagd naar het belang en de vormgeving van technisch lezen (klassikaal technisch lezen, technisch lezen in niveaugroepen binnen de klas of over de klassegrenzen heen), de gebruikte methoden en materialen, en de tijd die aan dit aspect van lezen per week wordt besteed. Uit de antwoorden bleek dat een opmerkelijk verschil tussen beide groepen van scholen was gelegen in het werken met niveaugroepen binnen de klas en over de klassegrenzen heen. Verder waren er geen systematische verschillen tussen beide groepen wat betreft de aandacht voor en de vormgeving van technisch lezen.

Het derde onderdeel van de vragenlijst heeft betrekking op begrijpend lezen. Dit aspect van het lezen wordt globaal omschreven als het kunnen begrijpen, verklaren en kritisch beoordelen van de informatie uit de tekst. In verband met het onderzoek werden over dit aspect van lezen de meeste, merendeels geprecodeerde, vragen gesteld. Gevraagd werd naar het belang en de toepassing van verschillende leesvormen zoals klassikaal begrijpend lezen, groepslezen, forumlezen en kringlezen, de gebruikte methoden en teksten, het gebruik van verschillende soorten van vragen (weet- en denkvragen), en de tijd die aan dit aspect van lezen per week wordt besteed. Opvallend was dat in beide groepen van scholen gemiddeld per week meer tijd aan technisch dan aan begrijpend lezen werd besteed. Uit de antwoorden bleek dat bij de toepassing van kringlezen sprake was van een duidelijk verschil tussen beide groepen. Verder waren er geen systematische verschillen tussen beide groepen wat betreft de aandacht voor en de vormgeving van begrijpend lezen.

Het vierde onderdeel van de vragenlijst heeft betrekking op studerend lezen. Dit aspect van lezen wordt globaal omschreven als het op een efficiënte

en systematische wijze verzamelen van informatie uit verschillende bronnen. Bij dit onderdeel werd gevraagd naar het belang en de toepassing van studerend lezen, de gebruikte methoden en teksten, eventuele oefeningen in het schematiseren en samenvatten van tekstuele informatie, en de tijd die aan dit aspect van lezen wordt besteed. Uit de antwoorden bleek dat de leerkrachten van de controlescholen gemiddeld meer tijd aan studerend lezen besteedden dan de leerkrachten van de experimentele scholen. Verder werden er geen systematische verschillen tussen beide groepen van scholen gevonden.

Het laatste onderdeel van de vragenlijst heeft betrekking op belangstelling voor lezen. Dit aspect van lezen wordt omschreven als (het ontwikkelen van) een positieve houding ten opzichte van lezen en leesmateriaal. Bij dit onderdeel werd gevraagd naar de toepassing van vrij lezen in de klas, de aard van het gebruikte leesmateriaal, het voorlezen van een boek, en de tijd die aan dit aspect van lezen per week wordt besteed. Uit de antwoorden bleek dat er sprake was van een gering verschil tussen beide groepen scholen wat betreft de tijd die aan de ontwikkeling van belangstelling voor het lezen werd besteed. Verder werden er geen systematische verschillen tussen beide groepen gevonden.

Zoals reeds eerder is opgemerkt, kan met behulp van de antwoorden op de betreffende vragenlijst niet meer dan een algemene indruk worden verkregen van het leesonderwijs in beide groepen van scholen. Een vergelijking van de antwoorden van de leerkrachten van de experimentele scholen met die van de controlescholen wijst er op dat er tussen beide groepen geen belangrijke systematische verschillen bestaan wat betreft de vormgeving van het leesonderwijs. De verschillen in antwoorden tussen de leerkrachten binnen elk van de groepen zijn meestal aanzienlijk groter dan de verschillen tussen de groepen.

4.7. Dataverzameling

In deze paragraaf wordt beschreven welke gegevens in het kader van dit onderzoek zijn verzameld en op welke wijze dat is gebeurd. Binnen het onderzoek wordt onderscheid gemaakt tussen een descriptief en een experimenteel gedeelte. De steekproef wordt om die reden verdeeld in een groep experimentele scholen en een groep controlescholen. Voor de dataverzameling is

dit onderscheid niet relevant. Hetgeen hier wordt beschreven, geldt voor zowel de experimentele scholen als de controlescholen.

4.7.1. Leerlinggegevens

Voor de testafnames is van elke leerling de naam, het geslacht en de geboortedatum genoteerd. De namen van de leerlingen zijn alleen gebruikt bij de testafnames. Daarna werden ze vervangen door een codenummer. Met behulp van de geboortedatum is de leeftijd van de leerlingen bepaald op 1 november 1978.

Gedurende de loop van het onderzoek werd gevraagd naar het beroep van de vader van de leerlingen. Dit gegeven is gebruikt om de leerlingen te classificeren wat betreft het sociaal milieu. Daarbij is de Beroepenklapper van het Instituut voor Toegepaste Sociologie in Nijmegen (Van Westerlaak, Kropman & Collaris, 1975) gehanteerd. De beroepen van de vaders van de leerlingen werden door de leerkrachten met de nodige terughoudendheid verstrekt. Na codering van deze gegevens en na een controle op deze codering werden deze gegevens vernietigd. Hetzelfde is na afloop van het onderzoek gedaan met de lijsten met leerlingennamen. Tenslotte zijn van elke leerling de rapportcijfers voor lezen, taal en rekenen van het kerst- en paasrapport van het schooljaar 1978-1979 verzameld. De problemen die zich bij de codering en de interpretatie daarvan voordoen, komen in hoofdstuk V aan de orde.

4.7.2. Testafnames

Tijdstippen van afname

In tabel 4.16 wordt schematisch weergegeven op welk moment in welke school een bepaalde test moest worden afgenomen. Tijdens de voormeting werden eerst enkele gestandaardiseerde tests afgenomen. Voor de Schriftelijke Opdrachten (SO) van Brus en Van Bergen (1973) werd de instructietijd geschat op ongeveer 10 minuten en de in de handleiding genoemde testtijd van 30 minuten aangehouden. Voor de Woordenschattest (WST) van Stijnen (1975) werd op een instructietijd van 10 minuten gerekend, terwijl de maximale afnameduur op 60 minuten werd gesteld. Bij de Stilleestest (SLT) van Bakker (1972) werd op een instructietijd van ongeveer een kwartier en een testtijd van maximaal 50 minuten gerekend. Na de bovengenoemde tests werd in alle

scholen de A-vorm van de Free Response Cloze Test (FRCT) en de Eén-Minuut-Test (EMT) van Brus en Voeten (1973) afgenomen. Voor de Free Response Cloze Test werd op een instructietijd van 10 minuten gerekend, terwijl als maximum testtijd 35 minuten werd aangehouden. Deze testtijd bleek tijdens eerder onderzoek ruim voldoende te zijn voor leerlingen van het vierde leerjaar. In de laatste weken van november werd de Standard Progressive Matrices van Raven (1958) en de A-vorm van de Multiple Choice Cloze Test (MCCT) afgenomen. Voor de test van Raven werd op 10 minuten instructietijd en maximaal 60 minuten testtijd gerekend. Voor de Multiple Choice Cloze Test werd 10 minuten instructietijd en maximaal 45 minuten testtijd aangehouden. Tijdens de laatste periode van de voormeting werden de A-vormen van de drie taakgerichte tests afgenomen nl. de Hoofdgedachtetest (HGD), de Redeneertest (RED) en de Bedoelingtest (BED). Elk van deze tests bestaat uit twee ongeveer gelijke delen. Het was de bedoeling dat beide delen van eenzelfde test in een klas op dezelfde dag werden afgenomen met een "pauze" tussen beide delen; bijvoorbeeld een deel vóór en een deel ná het speelkwartier, of een deel vóór de middagpauze en een deel erna. Voor het eerste deel van elke test werd een instructietijd van 10 minuten aangehouden. Voor het tweede deel werd de instructie in het kort herhaald. Vijf minuten werd daarvoor voldoende geacht. De afnameduur van de Hoofdgedachte- en de Redeneertest werd op 35 minuten per deel gesteld; de afnameduur van de Bedoelingtest was maximaal 40 minuten per testdeel.

Tijdens de nameting werden in een periode van vier weken de B-vormen van de vijf ontwikkelde tests afgenomen. De drie taakgerichte tests werden in de in tabel 4.16 aangegeven week op één dag in de klas afgenomen. De beide cloze tests werden beide in één week op verschillende dagen afgenomen. Voor de instructietijd en de maximale afnameduur werden dezelfde tijden aangehouden als in de voormeting.

Uit de tabel blijkt dat er verschillende afnamevolgordes per school zijn gehanteerd. De afnamevolgordes in de scholen 1 en 5, de scholen 2 en 6, de scholen 3 en 7 en de scholen 4 en 8 zijn gelijk. De scholen 1, 2, 3 en 4 zijn de experimentele scholen, de scholen 5, 6, 7 en 8 zijn de controle-scholen. Er is voor verschillende afnamevolgordes gekozen om een eventueel "groei-" of "vergeet-"effect van de experimentele programma's in de experimentele scholen zoveel mogelijk te spreiden over de verschillende tests.

Tabel 4.16. Opzet testafnames.

176

		tijdstip	schoolnummer							
			1	2	3	4	5	6	7	8
voormeting	1978	23-27 okt	SO	SO	SO	SO	SO	SO	SO	SO
		30- 3 nov	WST	SLT	SLT	WST	WST	SLT	SLT	WST
		6-10 nov	SLT	WST	WST	SLT	SLT	WST	WST	SLT
		13-17 nov	FRCT EMT	FRCT EMT	FRCT EMT	FRCT EMT	FRCT EMT	FRCT EMT	FRCT EMT	FRCT EMT
		20-24 nov	SPM	MCCT	MCCT	SPM	SPM	MCCT	MCCT	SPM
		27- 1 dec	MCCT	SPM	SPM	MCCT	MCCT	SPM	SPM	MCCT
		11-15 dec	HGD RED	HGD RED	HGD RED	HGD RED	HGD RED	HGD RED	HGD RED	HGD RED
		18-22 dec	BED	BED	BED	BED	BED	BED	BED	BED
nameting	1979	5- 9 mrt	HGD	RED	BED	FRCT MCCT	HGD	RED	BED	FRCT MCCT
		12-16 mrt	RED	BED	FRCT MCCT	HGD	RED	BED	FRCT MCCT	HGD
		19-23 mrt	BED	FRCT MCCT	HGD	RED	BED	FRCT MCCT	HGD	RED
		26-30 mrt	FRCT MCCT	HGD	RED	BED	FRCT MCCT	HGD	RED	BED
retentiemeting	1979	7-11 mei	HGD	RED	BED	FRCT MCCT	HGD	RED	BED	FRCT MCCT
		14-18 mei	RED	BED	FRCT MCCT	HGD	RED	BED	FRCT MCCT	HGD
		21-23 mei	BED	FRCT MCCT	HGD	RED	BED	FRCT MCCT	HGD	RED
		28- 1 juni	FRCT MCCT	HGD	RED	BED	FRCT MCCT	HGD	RED	BED
		4- 8 juni	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Tijdens de retentiemeting werden opnieuw de A-vormen van de vijf ontwikkelde tests afgenomen. Ook deze maal werden dezelfde maximum testtijden aangehouden. Doordat de scholen op verschillende tijdstippen een aantal vrije dagen hadden gepland, was het te voorzien dat het niet mogelijk zou zijn om de vijf tests in een periode van vier weken af te nemen. Aan de testinstructeurs werd daarom de mogelijkheid geboden zonnodig een week langer door te gaan. Wel moest dezelfde volgorde van afnames aangehouden worden.

Het grootste deel van de tests is inderdaad op de in tabel 4.16. aangegeven tijdstippen afgenomen. Het was echter niet te voorkomen dat in enkele gevallen van de planning moest worden afgeweken. Een ernstige afwijking van het geplande schema was dat een van de controlescholen (nummer 5 in het schema) niet in staat was aan de retentiemeting deel te nemen. Het gevolg daarvan was dat het aantal leerlingen binnen de verschillende analyses op de data van de retentiemeting noodzakelijkerwijs geringer is dan was gepland.

Testinstructeurs

De tests zijn afgenomen door acht student-assistenten. Een studente fungeerde als algemene reserve; zij nam tevens het correctiewerk voor haar rekening. De tests werden gedurende het hele onderzoek door dezelfde personen afgenomen; een student-assistent nam op een school alle tests af.

Voor en tijdens de eerste weken van de voormeting werden ter voorbereiding van de testinstructeurs vijf bijeenkomsten georganiseerd, waarbij elk van de tests zowel wat betreft de inhoud als de instructie werd besproken. Bovendien werden op deze bijeenkomsten vragen en opmerkingen over de testafnames van de afgelopen weken besproken.

Elke test was voorzien van een volledig uitgeschreven instructie. Voor de gestandaardiseerde tests was deze instructie afkomstig uit de betreffende handleiding. Voor de ontwikkelde tests waren handleidingen beschikbaar die in ongeveer dezelfde vorm ook al bruikbaar bleken bij de proefafnames en bij de afnames voor de constructie van de parallelvormen. Bestudering van deze handleidingen werd als een essentieel onderdeel gezien van de voorbereiding evenals het bestuderen van de test zelf.

Tijdens en nameting en de retentiemeting werden geen instructie-bijeen-

komsten meer georganiseerd. Individuele vragen en opmerkingen werden besproken en genoteerd wanneer de testmaterialen werden opgehaald of ingeleverd.

De afnames van de tests impliceerden een aantal taken voor de betreffende instructeurs zoals het maken van afspraken met de klasse-leerkrachten, het voorbereiden van de testafnames, het afnemen van de test op het afgesproken tijdstip, het noteren van de testtijd van elke leerling bij de inname van het antwoordblad en het noteren van de leerlingcode op het antwoordblad. Bij de vijf ontwikkelde tests en bij de Stilleestest konden de gegevens rechtstreeks van het antwoordblad geponst worden. Bij de Woordenschattest en de Standard Progressive Matrices dienden de testinstructeurs de antwoorden van de leerlingen op ponsformulieren over te nemen. De Schriftelijke Opdrachten en de Free Response Cloze Test dienden eerst nagekeken te worden. Dit correctiewerk werd verricht door de reserve-instructeur. Verder dienden de instructeurs na elke testafname een vragenformulier in te vullen. Dit formulier is in Bijlage 4.2. opgenomen. Tenslotte was het de taak van de instructeurs een zodanig contact met de leerkrachten te onderhouden dat de uitvoering van het onderzoek gedurende het gehele jaar mogelijk bleef. Voor zover daarmee het verder onderzoek niet beïnvloed kon worden, werden de resultaten aan de leerkrachten gerapporteerd. Na de voormeting werden door de projectmedewerkers de resultaten van de gestandaardiseerde tests gerapporteerd. Na de laatste afname werden alle resultaten van de vijf ontwikkelde tests gerapporteerd.

4.7.3. Verloop van de testafnames

Informatie omtrent het verloop van de testafnames is voornamelijk afkomstig van de vragenformulieren die na elke testafname door de testinstructeurs werden ingevuld. Het vragenformulier had onder andere het doel na te gaan onder welke omstandigheden de verschillende tests werden afgenomen en om de testinstructeurs de gelegenheid te bieden voor het vastleggen van op- en aanmerkingen. Bovendien had dit formulier tot doel te registreren welke leerlingen afwezig waren of de test niet afkregen. Tenslotte was dit formulier bedoeld om een indruk te krijgen van de tijd die nodig was om de verschillende tests af te nemen. Aan de hand van hetgeen op dit formulier is ingevuld, zal een overzicht worden gegeven van het verloop van de test-

Omstandigheden waaronder de tests zijn afgenomen

Op elk van de acht scholen zijn gedurende het gehele onderzoek de tests door dezelfde personen afgenomen. Aan deze opzet kleefte onderzoekstechnisch een nadeel nl. als er sprake is van instructeurseffecten dan raken die op deze wijze versmolten met mogelijke school- of klasse-effecten. Daar staat echter het praktische voordeel tegenover dat de scholen en de klassen voor het afnemen van tests slechts met één persoon contact behoeven te onderhouden. Een voordeel dat niet onderschat moet worden in een situatie waarin naast het afnemen van tests in de experimentele scholen ook nog gedurende vijf weken door vier verschillende instructeurs leeslessen worden gegeven.

Op een enkele uitzondering na bleef de klasse-leerkracht bij de testafnames aanwezig. Veel opmerkingen over de omstandigheden waaronder de tests werden afgenomen, werden door de instructeurs niet vermeld. De enkele gevallen waarbij sprake was van afwijkingen van de gewenste situatie hebben ongetwijfeld bijgedragen tot een grotere omvang van de meetfouten. Er werd echter geen aanleiding in gezien om de scores van klassen op een bepaalde test uit de dataset te verwijderen. In sommige gevallen is dit bij enkele leerlingen gebeurd; deze gevallen zijn op dezelfde wijze gecodeerd als leerlingen die tijdens een afname afwezig waren.

Tijdens de nameting en de retentiemeting werd, vooral aan het einde van de beide testperiodes, opgemerkt dat sommige leerlingen minder geconcentreerd werkten. Enkele instructeurs wezen er op dat de leerlingen in een steeds sneller tempo gingen werken. Vooral de instructeur in de school die in tabel 4.16 is aangeduid als school 2 maakte melding van dit verschijnsel. Dit nam volgens hem zodanige vormen aan dat hij er aan twijfelde of de tests nog wel serieus gemaakt werden. Voordat met de analyse van de data begonnen werd, is daarom eerst nagegaan hoe lang elke leerling aan elke test of testdeel had gewerkt. Bij de nameting en de retentiemeting bleek een belangrijk deel van de leerlingen van school 2 de delen van de Hoofdgedachtetest, de Bedoelingtest en de Redeneertest in extreem snelle tijden te hebben gemaakt. Omdat dit verschijnsel zich beperkte tot deze school werd besloten de gegevens van die school voor wat betreft de na- en de retentiemeting buiten beschouwing te laten. Helaas valt hiermee een deel van

het aantal leerlingen van de experimentele scholen uit. Voorkomen wordt echter dat de experimentele effecten in een bepaalde richting worden beïnvloed.

Op- en aanmerkingen over de tests en het testmateriaal

De test Schriftelijke Opdrachten werd volgens de opmerkingen van een aantal instructeurs erg moeilijk gevonden. De instructie bij de Stilleestest was volgens sommige instructeurs te ingewikkeld. Over de Woordenschattest, de Eén-Minuu-Test en de Standard Progressive Matrices werden geen of weinig opmerkingen gemaakt. Wat betreft de Free Response Cloze Test werd opgemerkt dat deze test uitnodigde tot afkijken. Bij de Multiple Choice Cloze Test werd opgemerkt dat een aantal items triviaal was en een paar items onmogelijk vanuit de tekst konden worden beantwoord. Bij de Hoofdgedachtetest vroegen enkele instructeurs zich af of de goede antwoorden op de vragen ook niet gekozen konden worden zonder de teksten te lezen. Over de Redeneertest werd opgemerkt dat de leerlingen zich nog al eens vergisten bij het invullen van het antwoordblad. Over de Bedoelingtest werd opgemerkt dat de teksten door de leerlingen als interessant werden ervaren. Bij de nameting en retentiemeting zijn bijna geen opmerkingen gemaakt over de testmaterialen. Eén opmerking die bij al deze tests op de beide meetmomenten werd gemaakt, had betrekking op de bekendheid van de instructie en de instructievoorbeelden bij de leerlingen.

Afwezigheid van leerlingen

Op de vragenformulieren voor de instructeurs werd steeds bijgehouden welke leerlingen bij een testafname afwezig waren. Deze gegevens hadden onder andere een functie bij de controle op het codeer- en ponswerk. Een enkele maal kon een leerling, die een test gemist had, deze alsnog inhalen als de instructeur een volgende dag in een andere klas dezelfde test afnam. Met name in de nameting en retentiemeting is op deze mogelijkheid gewezen. Van de mogelijkheid om buiten de geplande periodes alsnog te proberen om door middel van individuele afnames de set van verzamelde gegevens te completeren is bewust afgezien. Daardoor kon niet worden voorkomen dat steeds een aantal leerlingen een bepaalde test gemist heeft. Het aantal leerlingen dat een test geheel of gedeeltelijk niet kon maken of waarvan de gegevens om andere redenen niet bruikbaar waren (buitenlandse kinderen), is

weergegeven in tabel 4.17.

Tabel 4.17. Ontbrekende gegevens per test.

	voormeting		nameting ^x		retentiemeting ^{xx}	
	abs.	perc.	abs.	perc.	abs.	perc.
Hoofdgedachtetest	33	6	22	5	12	3
Redeneertest	34	6	22	2	17	4
Bedoelingtest	43	8	14	3	11	3
FRCT	11	2	21	4	17	4
MCCT	25	5	28	6	14	3
Eén-Minuut-Test	11	2				
Standard Pr. Matrices	11	2				
Stilleestest	14	3				
Woordenschattest	28	5				
Schriftelijke Opdrachten	15	3				
Totaal aantal leerlingen	530		474		424	

^x zonder school 2

^{xx} zonder school 2 en 5

Het hoogste percentage ontbrekende gegevens treffen we aan bij de drie taakgerichte tests in de voormeting. In deze periode waren meer leerlingen wegens ziekte afwezig en kon niet zo veel gebruik worden gemaakt van de mogelijkheid de test alsnog te laten maken. De hoogte van de percentages is overigens niet onrustbarend. Bovendien kan worden aangenomen dat de geconstateerde uitval niet selectief is. De kinderen wisten nl. niet wanneer een test werd afgenomen.

Maximale testtijd

Voor elke test werd een maximale testtijd vastgesteld. Voor de gestandaardiseerde tests werd deze bepaald op basis van de gegevens uit de handleidingen, voor de ontwikkelde tests op basis van ervaringen uit eerdere

afnames. Tabel 4.18. geeft een overzicht van het aantal leerlingen dat een test niet binnen deze maximale tijd afkreeg.

Tabel 4.18. Aantal leerlingen dat de tests niet afmaakte in de maximale testtijd.

	voormeting		nameting ^x		retentiemeting ^{xx}	
	abs.	perc.	abs.	perc.	abs.	perc.
Hoofdgedachtetest (per deel)	16	1	2	0	0	0
Redeneertest (per deel)	7	1	3	1	0	0
Bedoelingtest (per deel)	15	1	2	0	0	0
FRCT	11	2	8	2	1	0
MCCT	67	13	15	3	4	1
Eén-Minuut-Test	n.v.t.					
Standard Pr. Matrices	0	0				
Stilleestest	7	1				
Woordenschattest	0	0				
Schriftelijke Opdrachten	114	22				
Totaal aantal leerlingen	530		474		424	

^x zonder school 2

^{xx} zonder school 2 en 5

Uit deze tabel blijkt dat voor bijna alle leerlingen de testtijd ruim genoeg was gekozen. Voor twee tests moet echter een uitzondering worden gemaakt nl. voor de Multiple Choice Cloze Test in de voormeting en voor de test Schriftelijke Opdrachten. In totaal kreeg 13% van de leerlingen de MCCT in de voormeting niet af binnen de maximale testtijd van 45 minuten. Tijdens de nameting en de retentiemeting is dit percentage veel geringer. Een vijfde deel van de leerlingen kreeg de test Schriftelijke Opdrachten niet binnen een half uur, de maximale tijd volgens de handleiding, af. Bij de analyse van de gegevens zal geen onderscheid worden gemaakt tussen niet gemaakte items en foutieve antwoorden, hetgeen betekent dat niet gemaakte items fout

worden gerekend. Is de test (of het testdeel) voor niet meer dan de helft gemaakt dan wordt gehandeld alsof de leerling afwezig was tijdens de afname. Dit was meestal het geval als leerlingen tijdens de testafname genoodzaakt waren voortijdig op te houden.

Samenvattend kunnen we vaststellen dat de dataverzameling grotendeels volgens plan is verlopen. Van de planning zoals die in tabel 4.16. is weer-gegeven, behoefde slechts enkele malen te worden afgeweken. Het percentage ontbrekende gegevens per test is binnen redelijke grenzen gebleven en er zijn geen aanwijzingen om aan te nemen dat de uitval selectief is geweest. Twee omstandigheden beïnvloeden echter tot op zekere hoogte de analysemogelijkheden. In de eerste plaats de uitval van een controleschool in de retentiemeting. In de tweede plaats de noodzaak de gegevens van een experimentele school voor wat betreft de nameting en de retentiemeting buiten beschouwing te laten, omdat getwijfeld werd aan de betrouwbaarheid van een deel van de gegevens. Deze omstandigheden hebben tot gevolg dat in de meeste analyses met kleinere steekproeven genoeg moet worden genomen dan aanvankelijk verwacht werd.

4.8. Voorbereiding en uitvoering van het experiment

In deze paragraaf geven we een beschrijving van de voorbereiding en uitvoering van het experiment dat op vier scholen is uitgevoerd. Uit het vervolg zal blijken dat een groot aantal activiteiten moest worden verricht om het experiment volgens plan te laten verlopen.

4.8.1. Selectie van het controleprogramma

Zoals uit de opzet van het onderzoek blijkt, was het de bedoeling om op de experimentele scholen vier verschillende programma's aan te bieden nl. drie experimentele programma's en een programma voor de controlegroep. Het laatste programma wordt aangeduid met de term controleprogramma. Alvorens het controleprogramma te selecteren, werden enkele criteria opgesteld waaraan het programma zou moeten voldoen. Deze criteria waren:

- a. het programma moet betrekking hebben op begrijpend lezen;
- b. het programma moet precieze aanwijzingen bevatten omtrent de doelstellingen van de lessen en de uit te voeren onderwijs- en leeractiviteiten;

- c. het programma moet nieuw zijn in die zin dat het niet op de experimentele en controlescholen wordt gebruikt.
- d. het programma mag geen oefeningen en opdrachten bevatten die in de drie experimentele programma's voorkomen;
- e. het programma moet gemakkelijk uitvoerbaar zijn en geen al te hoge eisen stellen aan de uitvoerders.

Na de inventarisatie van de leesprogramma's die op de experimentele en controlescholen werden gebruikt en na bestudering van een aantal Nederlandse leesprogramma's viel de keuze op het programma "Begrijpend lezen" dat door Bol, Köllen en Leistra (1979) is ontwikkeld voor de derde klas van de lagere school (S.V.O.-project 0275). In overleg met de ontwikkelaars van dit programma werd vastgesteld welke lessen in aanmerking zouden komen. In Bijlage 4.3. wordt weergegeven welke lessen uit het betreffende programma werden gekozen.

4.8.2. Selectie van de instructeurs

Voor de uitvoering van de programma's waren in totaal negen instructeurs nodig: acht instructeurs die de programma's zouden geven en één reserve-instructeur die bij ziekte of anderszins zou kunnen invallen. Besloten werd instructeurs aan te stellen die voldeden aan de volgende criteria:

- a. doctoraal student zijn in de onderwijskunde (pedagogiek), de taalkunde of psychologie;
- b. pedagogische academie als vooropleiding hebben;
- c. ondervinding hebben in het lager of buitengewoon lager onderwijs;
- d. in staat zijn leiding te geven aan een groep leerlingen van het vierde leerjaar;
- e. bereid zijn zich nauwkeurig aan de uitgewerkte lessen te houden (met andere woorden de lessen te geven zoals deze ontworpen waren);
- f. over voldoende tijd beschikken om op verschillende dagen in de week lessen te geven, deze voor te bereiden en te evalueren.

Mededelingen ad valvas gaven bekendheid aan de mogelijkheid om als instructeur/student-assistent aan het onderzoek mee te werken. Er werden voldoende gegadigden gevonden die voldeden aan de vooraf gestelde eisen. In totaal werden vijf studenten onderwijskunde, drie studenten Nederlands en één student orthopedagogiek geselecteerd.

4.8.3. Verdeling van de instructeurs over de scholen en de programma's

Zoals uit de opzet van het onderzoek blijkt, was het de bedoeling om in elke experimentele school vier groepen van leerlingen te vormen die elk op hetzelfde uur een verschillend programma zouden ontvangen. Elk programma zou - gegeven de vier experimentele scholen - dus vier keer worden aangeboden. Bij de opzet van het onderzoek werd besloten, dat elk programma door vier verschillende instructeurs zou worden gegeven. Zo zou bijvoorbeeld het programma Hoofdgedachte op vier verschillende scholen door vier verschillende instructeurs worden gegeven (om misverstand te voorkomen: de leerlingen kregen steeds les van dezelfde instructeur). Iedere instructeur zou twee verschillende programma's geven.

Na de selectie van de instructeurs werd vastgesteld welke instructeur welke combinatie van twee programma's zou geven. Bij vier programma's zijn er zes mogelijke combinaties van twee programma's. Bijvoorbeeld: een instructeur geeft het programma Hoofdgedachte en Bedoeling, een andere instructeur geeft Hoofdgedachte en Redeneren, etc. Omdat er acht instructeurs waren en zes mogelijke combinaties van programma's, moesten twee combinaties van programma's dubbel voorkomen. Besloten werd dat de combinatie Hoofdgedachte - Redeneren en Bedoeling - Controleprogramma dubbel zouden voorkomen.

De verdeling van de instructeurs over de scholen was zodanig, dat een vast team van vier instructeurs les zou geven op twee experimentele scholen, en een ander vast team van vier instructeurs op de andere twee experimentele scholen. De verdeling van de acht instructeurs (1-8) over de vier scholen (a-d) en de vier programma's (H, B, R, C) is weergegeven in figuur 4.2.

Restte nog te bepalen welke instructeur "instructeur 1" zou zijn (en dus het programma Bedoeling op school a zou geven en Redeneren op school b), welke instructeur "instructeur 2" zou zijn, etc. De toekenning van deze instructeurnummers gebeurde random.

Figuur 4.2. Verdeling van instructeurs over scholen en programma's.

		School				
		a	b	c	d	
Programma	H	2	4	6	5	instructeurs: 1 t/m 8 programma's H = Hoofdgedachte B = Bedoeling R = Redeneren C = Controle
	B	1	3	7	6	
	R	4	1	5	8	
	C	3	2	8	7	

4.8.4. Vorbereiding van de instructeurs

Voordat het experiment begon, werden in totaal vijf bijeenkomsten met de instructeurs gehouden. Tijdens de eerste bijeenkomst werden de instructeurs geïnformeerd omtrent het doel en de opzet van het onderzoek, de stand van zaken op dat moment en hun taak als instructeur. Tijdens de tweede bijeenkomst werd de verdeling van de instructeurs over de vier scholen en de vier programma's bekend gemaakt evenals de data en de tijden waarop de lessen gegeven moesten worden. Bovendien werd de nodige informatie verstrekt over de scholen, het aantal leerkrachten en leerlingen en dergelijke. Bij de twee volgende bijeenkomsten werden de instructeurs in groepen verdeeld en uitvoerig geïnformeerd over de bedoeling, de opbouw en de ontwikkeling van die programma's, die ze moesten geven. De reserve-instructeur woonde alle bijeenkomsten bij. Deze bijeenkomsten werden geleid door die medewerkers van het project, die het betreffende programma hadden ontwikkeld of op de hoogte waren van de ontwikkeling van het programma. Dit laatste geldt nl. voor het programma "Begrijpend lezen", i.c. het controleprogramma. Opge-merkt moet worden dat de instructeurs onkundig werden gehouden omtrent de functie van dit programma in het onderzoek.

De laatste bijeenkomst vóór het experiment was bestemd voor alle in-

structeurs. Tijdens deze bijeenkomst werden aanwijzingen gegeven omtrent de uitvoering van twee leeslessen die aan de vier programma's voorafgingen. Deze lessen waren afkomstig uit de herziene methode van Kohnstamm getiteld "Stillezen en er wat mee doen" (Bossers, 1975). Deze introductielessen waren bedoeld om de instructeurs en de leerlingen aan elkaar en aan de nieuwe situatie te laten wennen en om na te gaan of de organisatie van het geheel vlot zou verlopen. Tijdens de bijeenkomst werden tevens aanwijzingen gegeven omtrent de voorbereiding van de lessen, het gebruik van het bord, de omgang met de leerkrachten, het gebruik van de handleiding, het optreden in de klas, de manier van lesgeven, het inleveren van de lesmaterialen, het invullen van het evaluatieformulier, de taak van de coördinator en de wijze van handelen bij afwezigheid als gevolg van ziekte of anderszins. Aan het einde van de bijeenkomst werden de data en de tijden vastgesteld waarop tijdens het experiment bijeenkomsten werden gehouden ter voorbereiding en nabespreking van de lessen.

4.8.5. Voorbereiding van de leerkrachten

Zoals reeds eerder vermeld, werd na de selectie van de acht scholen met de hoofden en de leerkrachten van de vierde klas contact opgenomen en het doel en de opzet van het onderzoek besproken. Tijdens de voormeting werd met de leerkrachten van de experimentele scholen overleg gepleegd over vooral de organisatorische aspecten van het experiment. Voor elke experimentele school werden de data en de tijden vastgesteld waarop de lessen konden worden gegeven. Bovendien werd afgesproken welke vier lokalen dan gebruikt konden worden. De uitvoering van de programma's voor vier groepen leerlingen van één school moest nl. op hetzelfde uur plaats vinden. Vóór het begin van het experiment werden deze afspraken schriftelijk bevestigd. Om ongewenst oefenen tijdens of na het experiment te voorkomen, kon aan de leerkrachten geen informatie worden verstrekt over de inhoud van de verschillende programma's en was het voor hen niet mogelijk de lessen bij te wonen. Tijdens het overleg met de leerkrachten werd - opnieuw - voor deze situatie begrip gevraagd en ook getoond.

De leerkrachten van de experimentele scholen werden niet alleen geïnformeerd over de taken van de instructeurs, maar ook over de rol van de coördinator. Tijdens de uitvoering van de lessen zou nl. op elke experimentele

school een coördinator aanwezig zijn. De coördinator, een medewerker van het project, had tot taak eventuele moeilijkheden in het organisatorische vlak op te lossen. Hij droeg de zorg voor het tijdig beginnen en eindigen van de lessen, hij controleerde of de leslokalen voor gebruik gereed waren en na afloop weer netjes werden afgeleverd; hij zorgde dat de leerlingen in het juiste lokaal terecht kwamen, onderhield het contact met de leerkrachten en het hoofd van de school en kon in noodgevallen, als de reserve-instructeur niet meer tijdig kon worden opgeroepen, zelf inspringen.

4.8.6. Verdeling van de leerlingen over de programma's

Binnen elke experimentele school werden de leerlingen van de vierde klassen in vier groepen verdeeld. De verdeling van de leerlingen in groepen verliep als volgt. Voor iedere leerling werd een factorscore berekend. Deze factor was de eerste factor van een principale componenten-analyse (met enen op de diagonaal), berekend over alle leerlingen van de experimentele scholen en de controlescholen tezamen over gegevens verzameld tijdens de voormeting. Bij deze principale componenten-analyse werden de volgende tests betrokken: Hoofdgedachtetest, Bedoelingtest, Redeneertest, MCCT, FRCT, Stilleestest van Bakker (1972) en Schriftelijke Opdrachten van Brus en Van Bergen (1973). Als extractie criterium werd een eigenwaarde groter of gelijk aan 1 gehanteerd; bovendien werd het aantal te trekken factoren niet gefixeerd. Aangenomen werd dat de score op de eerste factor een maat vormde voor leesbekwaamheid.

Wat betreft de berekening van de factorscores is de volgende informatie van belang. De eerste factor van de principale componenten-analyse had een eigenwaarde van 4.56 en verklaarde 65.1% van de variantie; de tweede factor had een eigenwaarde van .68 en verklaarde 9.7% van de variantie. De ladingen van de tests op de eerste factor worden in tabel 4.19. weergegeven.

Tabel 4.19. Ladingen van de tests op de eerste factor (N = ca. 500).

	Factor 1
Hoofdgedachtetest	.82
Bedoelingtest	.85
Redeneertest	.72
MCCT	.79
FRCT	.81
Stilleestest	.79
Schriftelijke Opdrachten	.85

De factorscore is berekend volgens de formule:

$$f_j = \sum_{i=1}^7 f_{ci} z_{ij} \quad (1)$$

waarin f_j de factorscore voor persoon j is,

f_{ci} de factorscore-coëfficiënten voor de 7 (i) tests zijn en

z_{ij} de gestandaardiseerde score is van persoon j op test i .

De factorscore-coëfficiënten zijn berekend volgens de formule (in matrix-rotatie): $(L'L)^{-1} L'$. (De inverse van het product van de getransponeerde vector van ladingen (L) met de vector van ladingen vermenigvuldigd met de getransponeerde vector van ladingen). Dit leverde een factorscore-coëfficiënt vector op. In tabel 4.20. worden de factorscore-coëfficiënten voor de verschillende tests weergegeven.

Tabel 4.20. Factorscore-coëfficiënten voor de tests.

	Factorscore-coëfficiënt
Hoofdgedachtetest	.18092
Bedoelingtest	.18605
Redeneertest	.15790
MCCT	.17331
FRCT	.17871
Stilleestest	.17428
Schriftelijke Opdrachten	.18664

Uit deze coëfficiënten blijkt dat de gewichten waarmee elk van de gestandaardiseerde testscores gewogen worden niet zo veel van elkaar verschillen. Het laagste gewicht heeft de Redeneertest, het hoogste de Schriftelijke Opdrachten.

Bij de berekening deed zich het probleem voor dat een aantal leerlingen één of meerdere test(s) niet had gemaakt. Het aantal leerlingen (van alle scholen) dat een bepaalde test niet had gemaakt, wordt in tabel 4.21. weerge-

gegeven.

Tabel 4.21. Aantal leerlingen dat één of meerdere test(s) miste.

	Aantal leerlingen
Hoofdgedachtetest	36
Bedoelingtest	53
Redeneertest	37
MCCT	28
FRCT	13
Stilleestest	15
Schriftelijke Opdrachten	15

Omdat het voor de matching gewenst was dat alle leerlingen, met name de leerlingen van de experimentele scholen, een factorscore hadden waarop ze ingedeeld konden worden, is voor de leerlingen die één of meer test(s) gemist hadden formule (1) zodanig aangepast dat ook in dat geval een factorscore berekend kon worden. De aangepaste formule (1) ziet er dan als volgt uit:

$$f_j = \frac{t}{t-n_j} \sum_i f_{c_i} z_{ij} \quad (2)$$

waarbij nu gesommeerd wordt over het aantal variabelen waarvoor gegevens beschikbaar zijn en waarin

t het totale aantal tests (7) is en

n_j het aantal tests is dat door leerling j gemist is.

Op deze manier zou eventueel nog een factorscore berekend kunnen worden als zes van de zeven testuitslagen ontbreken. Voor de leerlingen die aan het experiment deelnamen, deed zich dit extreme geval echter niet voor.

Op grond van hun factorscore werden de leerlingen van elke experimentele school nu geordend van hoog naar laag. Vervolgens werden groepen van telkens vier leerlingen gevormd. De vier leerlingen van een school met de hoogste factorscores vormden een viertal, de vier leerlingen met de daarop volgende factorscores vormden het tweede viertal, etc. Tenslotte werden de leerlingen van elk viertal op random wijze verdeeld over de vier programma's. Zo kreeg van de vier leerlingen met de hoogste factorscores één leerling het programma Hoofdgedachte, één leerling het programma Bedoeling, één leerling het programma Redeneren en één leerling het Controleprogramma.

4.8.7. Voorbereiding en uitvoering van de lessen

In de week van 15 januari 1979 werden de twee lessen uit de methode "Stillezen en er wat mee doen" gegeven. Op één school moest de reserve-instructeur een les overnemen wegens ziekte van een instructeur. De introductielessen verliepen verder volgens plan. De instructeurs en de leerlingen leerden elkaar kennen en raakten gewend aan de nieuwe situatie; de organisatie leverde geen problemen op.

In de periode van 15 januari tot 24 februari 1979 werden wekelijks bijeenkomsten gehouden, waarbij de instructeurs opnieuw in groepen werden verdeeld. Elke instructeur woonde twee bijeenkomsten bij: één bijeenkomst voor bijvoorbeeld het programma Hoofdgedachte en één bijeenkomst voor het programma Bedoeling. De reserve-instructeur volgde wekelijks alle bijeenkomsten. Deze bijeenkomsten waren zodanig gepland, dat telkens de twee lessen besproken werden, die de volgende week gegeven moesten worden. Vóór elke bijeenkomst bestudeerden de instructeurs de handleiding en het leerlingenmateriaal van twee lessen. Tijdens deze wekelijkse bijeenkomsten werden niet alleen de volgende lessen voorbereid, maar konden ook problemen worden besproken die zich bij de al gegeven lessen hadden voorgedaan. Het zal duidelijk zijn dat deze bijeenkomsten een zekere controle op de voorbereiding en uitvoering van de lessen mogelijk maakten.

De uitvoering van de vier programma's van elk tien lessen vond in de periode van 22 januari tot 24 februari 1979 plaats. Elke groep leerlingen (in totaal 16 groepen) kreeg per week twee lessen van elk \pm 45 minuten. Eén instructeur gaf per week twee lessen van het ene programma aan een

groep leerlingen van de ene school en twee lessen van het andere programma van een andere school. Eén instructeur gaf gedurende vijf weken 20 lessen; in totaal werden dus 160 lessen gegeven. Wat betreft de uitvoering van deze lessen kan worden opgemerkt dat bijna alle lessen op de geplande data en tijden werden gegeven. Op één dag konden wegens barre weersomstandigheden de lessen op twee experimentele scholen geen doorgang vinden. De lessen werden kort daarna gegeven. Wegens ziekte werden in totaal 3 lessen (van de 160) overgenomen door de reserve-instructeur. De handleiding en het leerlingmateriaal van elke les werden tijdig ingeleverd evenals de evaluatieformulieren.

4.8.8. Evaluatie van de lessen en programma's

Elke instructeur vulde zo spoedig mogelijk na afloop van een les een evaluatieformulier in en overhandigde dit tijdens de wekelijkse bijeenkomsten aan de betreffende projectmedewerkers. Dit formulier is opgenomen in Bijlage 4.4. Na afloop van de programma's vulden de instructeurs een evaluatieformulier in betreffende het gehele programma. Dit formulier is opgenomen in Bijlage 4.5.

Zoals uit het evaluatieformulier voor elke les blijkt, moeten bij de eerste twee vragen de namen (en codenummers) van de leerlingen worden genoemd, die de les niet of slechts gedeeltelijk hebben gevolgd. Deze informatie is van belang in verband met een mogelijke uitval van leerlingen. De volgende vragen hebben betrekking op de duur van de les, het al of niet bereiken van het doel of de doelen van de les, de motivatie van de leerlingen tijdens de les etc. De laatste vraag heeft uitsluitend een administratieve functie. Zoals uit de globale formulering van een aantal vragen reeds blijkt, heeft het evaluatieformulier voor elke les slechts een beperkte functie nl. een globale registratie van het verloop van de lessen. Gezien het geringe aantal instructeurs voor elk programma mag aan de antwoorden op de vragen slechts een beperkte waarde worden toegekend.

Het evaluatieformulier betreffende het gehele programma bevat vragen omtrent het belang van het programma, de moeilijkheidsgraad, de structuur of opbouw van het programma, de aanwijzingen voor de instructeur, het leerlingmateriaal etc. Deze vragen zijn evenals een aantal vragen van het evaluatieformulier voor elke les tamelijk globaal van aard en hebben slechts een

bepaalde functie nl. inventariseren wat vier instructeurs van een programma denken.

In het volgende zullen we voor elk programma de conclusie vermelden, die uit de antwoorden van de instructeurs op beide evaluatieformulieren kan worden afgeleid, evenals de antwoorden op enkele belangrijke vragen. Voor meer informatie verwijzen we naar Deelrapport 5 (Aarnoutse et al., 1980).

Programma Hoofdgedachte

Uit de antwoorden van de instructeurs kan worden afgeleid dat de lessen in het algemeen volgens plan zijn uitgevoerd en weinig problemen hebben opgeroepen. Over het gehele programma zijn de instructeurs over het algemeen tevreden; één instructeur heeft moeite met de strakke opzet van het programma.

Volgens de instructeurs werd het doel (of de doelen) van het merendeel van de lessen bereikt en was de motivatie van de meeste leerlingen tijdens de tien lessen goed of ruim voldoende. Uit de antwoorden van drie instructeurs blijkt dat de motivatie na de zesde les enigszins afnam. Volgens de instructeurs werd het merendeel van de geplande activiteiten of oefeningen uitgevoerd en verliepen de meeste activiteiten of oefeningen goed. Alle instructeurs zijn van mening dat het programma "niet moeilijk en niet gemakkelijk" is voor leerlingen van het vierde leerjaar.

Programma Redeneren

Uit de antwoorden van de vier instructeurs kan worden afgeleid dat de lessen in het algemeen volgens plan zijn uitgevoerd en weinig problemen hebben opgeroepen. Over het gehele programma zijn de instructeurs over het algemeen tevreden.

Volgens de instructeurs werd het doel (of de doelen) van het merendeel van de lessen bereikt en was de motivatie van de meeste leerlingen tijdens de tien lessen goed of ruim voldoende. Een instructeur vond de motivatie van de leerlingen bij vier lessen zwak. Volgens de opgave van de instructeurs werd het merendeel van de geplande activiteiten of oefeningen uitgevoerd en verliepen de meeste activiteiten of oefeningen goed. Drie instructeurs vonden het programma "niet moeilijk en niet gemakkelijk" voor leerlingen van de vierde klas. Een instructeur vond sommige lessen moeilijk en

andere gemakkelijk voor deze leerlingen.

Programma Bedoeling

Uit de antwoorden van de instructeurs kan worden afgeleid dat de lessen in het algemeen volgens plan zijn uitgevoerd, al vormden de moeilijkheidsgraad en de overladenheid van een aantal lessen een complicerende factor. De instructeurs zijn over het programma niet allen even tevreden.

Volgens twee instructeurs werd het doel (of de doelen) van het merendeel van de lessen bereikt; twee andere instructeurs zijn minder positief en weten niet of het doel bereikt werd door de leerlingen. Volgens de instructeurs was de motivatie van de leerlingen tijdens de meeste lessen ruim voldoende of voldoende. Het merendeel van de geplande activiteiten of oefeningen werd volgens de instructeurs uitgevoerd en verliep goed. Een aantal oefeningen kon echter wegens tijdgebrek of overladenheid niet aan de orde komen. Twee instructeurs vonden het programma moeilijk voor leerlingen van de vierde klas; de twee andere instructeurs zijn van mening dat de moeilijkheidsgraad van de lessen varieerde van gemakkelijk tot moeilijk.

Controleprogramma

Uit de antwoorden van de instructeurs op vragen van de beide evaluatieformulieren kan worden afgeleid dat de lessen in grote lijnen volgens plan zijn uitgevoerd. Over het hele programma zijn de instructeurs over het algemeen niet zo tevreden.

Volgens de instructeurs werd het doel (of de doelen) van het merendeel van de lessen bereikt en was de motivatie van de leerlingen tijdens de meeste lessen ruim voldoende. Volgens de opgave van de instructeurs werd het merendeel van de geplande activiteiten of oefeningen uitgevoerd en verliepen de meeste activiteiten goed. Drie instructeurs vonden het programma gemakkelijk voor leerlingen van het vierde leerjaar; één instructeur vond het programma niet moeilijk of gemakkelijk. Over de structuur of opbouw van het programma en over de aanwijzingen voor de leerkracht (de handleiding) zijn de instructeurs niet tevreden.

Dit hoofdstuk bevat twee paragrafen. In par. 5.1. worden eerst de vraagstellingen van het descriptieve gedeelte van het onderzoek verder gepreciseerd en bij een aantal vragen verwachtingen geformuleerd. Daarna worden de vraagstellingen c.q. verwachtingen in een aantal subparagrafen beantwoord respectievelijk onderzocht. Op het einde van de paragraaf volgt een samenvatting van de belangrijkste resultaten. In par. 5.2. worden de vraagstellingen van het experimentele gedeelte van het onderzoek nader omschreven in de vorm van hypothesen. Na een beschrijving van de analyse van de data volgt een samenvatting.

5.1. Het descriptieve gedeelte van het onderzoek

5.1.1. Vraagstellingen en verwachtingen

Zoals uit hoofdstuk IV blijkt, hebben de vraagstellingen van het descriptieve gedeelte van het onderzoek betrekking op de prestaties van leerlingen van het vierde leerjaar op de ontwikkelde tests, de betrouwbaarheid van de tests, de validiteit van de tests en de verandering in prestaties over drie meetmomenten. De vragen kunnen als volgt worden gepreciseerd:

- a. Welke prestaties leveren leerlingen van het vierde leerjaar op de ontwikkelde tests tijdens de voormeting?
 - In hoeverre verschillen de prestaties van de betreffende onderzoeksscholen van prestaties van een aselechte steekproef van scholen?
 - Welke verschillen in prestaties bestaan er tussen sociale milieus?
 - Welke verschillen in prestaties bestaan er tussen de scholen?
- b. Hoe betrouwbaar zijn de ontwikkelde tests?
 - Welke waarde hebben de coëfficiënten van stabiliteit, a-specificiteit en interne consistentie?
 - Treden er verschillen in betrouwbaarheidscoëfficiënten op tussen experimentele scholen en controlescholen?
- c. Hoe valide zijn de ontwikkelde tests?
 - In hoeverre zijn de ontwikkelde tests homogeen?
 - Welke samenhang bestaat er tussen de ontwikkelde tests?

- Welke samenhang bestaat er tussen de ontwikkelde tests en een gestandaardiseerde test voor technisch lezen?
 - Welke samenhang bestaat er tussen de ontwikkelde tests en gestandaardiseerde tests voor begrijpend lezen?
 - Welke samenhang bestaat er tussen de ontwikkelde tests en een gestandaardiseerde test voor woordenschat?
 - Welke samenhang bestaat er tussen de ontwikkelde tests en een gestandaardiseerde test voor non-verbale intelligentie?
 - Welke samenhang bestaat er tussen de scores op de ontwikkelde tests en cijfers voor lezen, taal en rekenen?
 - In hoeverre bestaat er ruimte voor de veronderstelling dat de taakgerichte tests een specifiek aspect van begrijpend lezen meten?
- d. Wat is de verandering in prestaties op de ontwikkelde tests wanneer geen experimentele programma's gegeven worden?
- Hoe groot is de gemiddelde verandering tijdens de totale periode en tijdens de beide te onderscheiden deelperiodes (voormeting-nameting en nameting-retentiemeting)?
 - Hoe groot zijn de individuele verschillen in verandering tijdens de totale periode en tijdens de beide deelperiodes?
 - Welke factoren hangen samen met individuele verschillen in verandering?

In hoofdstuk III hebben we voor een aantal vragen genoemd onder c. verwachtingen geformuleerd die voor een groot deel zijn gebaseerd op de bestudeerde onderzoeksliteratuur. Wat betreft de convergente en discriminante validiteit van de ontwikkelde tests verwachten we:

- a. een aanzienlijke samenhang tussen de vijf ontwikkelde tests;
- b. een vrij lage samenhang tussen de taakgerichte tests en een maat voor technisch lezen en waarschijnlijk een middelmatige samenhang tussen de cloze tests en deze maat voor technisch lezen;
- c. een aanzienlijke samenhang tussen de vijf ontwikkelde tests en een maat voor begrijpend lezen in het algemeen;
- d. een aanzienlijke samenhang tussen de vijf ontwikkelde tests en een maat voor woordenschat;

e. een middelmatige tot vrij lage samenhang tussen de ontwikkelde tests en een maat voor non-verbale intelligentie.

5.1.2. De prestaties van de leerlingen op de ontwikkelde tests bij de voormeting

Frequentieverdelingen

Voordat de gegevens betreffende de prestaties van de leerlingen gepresenteerd worden, zal worden nagegaan in hoeverre de prestaties van de leerlingen van de onderzoeksscholen verschillen van de prestaties van een aselechte steekproef van scholen.

Uit de procedure die gehanteerd is bij de selectie van de acht onderzoeksscholen volgt dat de gegevens van de onderzoeksscholen alleen generaliseerbaar zijn naar de verzameling "grote scholen voor gewoon lager onderwijs van het voormalig stadsgewest Nijmegen, inclusief de gemeente Elst". Toch is het mogelijk een indruk te krijgen van de landelijke prestaties op de vijf ontwikkelde tests door de acht onderzoeksscholen op bepaalde variabelen te vergelijken met een steekproef die "representatief" kan worden geacht voor de landelijke populatie scholen voor gewoon lager onderwijs.

Gegevens betreffende een steekproef uit de landelijke populatie scholen voor gewoon lager onderwijs zijn beschikbaar van een ijkingsonderzoek dat in het schooljaar 1970-1971 is uitgevoerd door medewerkers van het Instituut voor Onderwijskunde te Nijmegen. In dit onderzoek werden de volgende tests genormeerd: de Eén-Minuut-Test van Brus en Voeten (1973), de Stilleestest van Bakker (1972), de test Schriftelijke Opdrachten van Brus en Van Bergen (1973), de Woordenschattest van Stijnen (1975) en de Toetsen voor Inzichtelijk Rekenen van Mommers en Smits (1973). In het kader van het onderzoek naar de validiteit van deze tests werden nog enkele andere tests afgenomen. Bij het onderzoek waren de leerjaren 2 tot en met 6 betrokken. De ijkingssteekproef bestond uit 25 scholen voor gewoon lager onderwijs; de steekproef was niet gestratificeerd. In verband met het geringe aantal random geselecteerde scholen en de uitval van enkele scholen werd een controle van de "representativiteit" van de uiteindelijke steekproef op een aantal relevante variabelen uitgevoerd. Geconcludeerd werd dat de steekproef voor de

variabelen waarvoor gecontroleerd kon worden, zeer redelijk aan de eis van "representativiteit" voldeed.

Uit het voorgaande blijkt dat een aantal tests uit het bovengenoemde ijkingsonderzoek ook in het onderhavige onderzoek wordt afgenomen nl. de Eén-Minuut-Test, de Stilleestest, de test Schriftelijke Opdrachten en de Woordenschattest. Zowel in het ijkingsonderzoek als in het huidige onderzoek werden de leerlingen ingedeeld naar sociaal milieu (SM) op basis van het beroep van de vader. Hierbij werd gebruik gemaakt van een klassificatiesysteem dat ontwikkeld is door het Instituut voor Toegepaste Sociologie te Nijmegen (Van Westerlaak, Kropman & Collaris, 1975). In dit systeem worden zes beroepsniveaus onderscheiden nl. 1 = ongeschoolde arbeid, 2 = geschoolde arbeid, 3 = lagere employees, 4 = kleine zelfstandigen, 5 = middelbare employees, 6 = hogere beroepen. Is het beroep van de vader onbekend, dan wordt dit aangegeven door de code 9.

In tabel 5.1. worden de prestaties weergegeven van de leerlingen uit het vierde leerjaar van de IJkingsgroep (1970-1971) op de Eén-Minuut-Test, de Woordenschattest en de test Schriftelijke Opdrachten én de prestaties van de leerlingen van de Onderzoeksgroep op dezelfde tests. In beide groepen zijn de leerlingen opgesplitst naar sociaal milieu. De Onderzoeksgroep wordt gevormd door de acht bij het huidige onderzoek betrokken scholen; de gegevens van deze groep hebben betrekking op de voormeting. Van de Stilleestest zijn voor de IJkingsgroep geen gegevens gepubliceerd; deze test wordt daarom niet in tabel 5.1. opgenomen.

Uit tabel 5.1. valt het volgende te constateren:

- a. De gemiddelde score van de Onderzoeksgroep op de drie tests is hoger dan van de IJkingsgroep. Het verschil in gemiddelde tussen beide groepen op de Eén-Minuut-Test, de Woordenschattest en de Schriftelijke Opdrachten is respectievelijk 3.8, 1.3 en 1.1. Uitgedrukt in de standaarddeviatie van de respectievelijke tests is het verschil in gemiddelde 0.27 s, 0.12 s en 0.16 s.
- b. De verschillen in spreiding vertonen geen consistent beeld. Bij de Eén-Minuut-Test is de spreiding in de Onderzoeksgroep groter, bij de Woordenschattest en de Schriftelijke Opdrachten kleiner.
- c. De frequenties waarmee de sociale milieus voorkomen, zijn in de Onderzoeksgroep en de IJkingsgroep niet gelijk. De lagere milieus 1 tot en

Tabel 5.1. Prestaties van de leerlingen van de Onderzoeksgroep en de IJkingsgroep, gespecificeerd naar sociaal milieu (SM).

EEN-MINUUT-TEST

		Onderzoeksgroep (okt/nov)				IJkingsgroep (feb/mrt)			
		aantal lln.		\bar{x}		s		\bar{x}	
		abs.	%						
SM 1	39	7.5	63.8	15.0		75	11.5	61.8	12.5
2	152	29.2	66.3	16.0		224	34.5	64.9	13.1
3	105	20.2	67.4	12.8		148	22.8	67.4	13.1
4	42	8.1	67.0	13.8		106	16.3	62.4	16.0
5	80	15.4	68.0	11.8		47	7.2	70.9	12.2
6	75	14.4	71.9	14.4		50	7.7	66.8	14.1
9	27	5.2	64.7	14.0		-	-	-	-
Totaal	520	100	67.4	14.3		650	100	63.6	14.2

WOORDENSCHATTEST

		Onderzoeksgroep (okt/nov)				IJkingsgroep (feb/mrt)			
		aantal lln.		\bar{x}		s		\bar{x}	
		abs.	%						
SM 1	36	7.1	32.0	10.1		-	12.3	32.3	10.6
2	150	29.8	34.4	10.2		-	32.0	33.8	11.1
3	101	20.0	37.8	9.9		-	20.7	40.2	10.0
4	42	8.3	37.4	8.4		-	17.4	36.4	10.5
5	80	15.9	41.1	9.1		-	6.5	40.2	10.8
6	72	14.3	44.5	8.4		-	6.2	41.8	11.0
9	23	4.6	33.7	9.4		-	4.9	-	-
Totaal	504	100	37.6	10.2		652	100	36.3	11.3

SCHRIFTELIJKE OPDRACHTEN

		Onderzoeksgroep (okt/nov)				IJkingsgroep (feb/mrt)			
		aantal lln.		\bar{x}		s		\bar{x}	
		abs.	%						
SM 1	40	7.7	16.3	6.1		74	11.7	15.4	6.9
2	148	28.6	17.9	6.9		217	34.3	16.5	6.9
3	103	19.9	19.0	6.7		146	23.1	19.8	6.7
4	42	8.1	18.6	6.4		105	16.6	17.9	6.9
5	81	15.6	20.3	6.2		43	6.8	20.7	5.8
6	77	14.9	22.0	5.7		48	7.6	20.3	6.6
9	27	5.2	15.2	6.2		-	-	-	-
Totaal	518	100	18.9	6.6		633	100	17.8	7.0

* De percentages leerlingen per SM bij de Woordenschattest zijn percentages van de leerlingen in leerjaar 3, 4, 5 en 6 samen. Afzonderlijke percentages voor leerjaar 4 zijn niet beschikbaar. De gemiddelde scores en de standaarddeviaties per SM zijn wel van leerjaar 4.

met 4 (met name 1 en 4) zijn in de Onderzoeksgroep ondervertegenwoordigd; de hogere milieus 5 en 6 zijn oververtegenwoordigd.

- d. De gemiddelde scores per sociaal milieu vertonen vaker een verschil ten gunste van de Onderzoeksgroep dan ten gunste van de IJkingsgroep. De verschillen in gemiddelde scores per sociaal milieu zijn alleen consistent over de drie tests voor de milieus 2 en 4 (beide ten gunste van de Onderzoeksgroep).

Het is duidelijk dat de Onderzoeksgroep hoger scoort dan de IJkingsgroep. De (hoger scorende) hogere milieus zijn in de Onderzoeksgroep oververtegenwoordigd. Bovendien bestaat binnen de sociale milieus de tendens dat de Onderzoeksgroep hoger scoort, ondanks het feit dat de tests bij de Onderzoeksgroep drie maanden eerder in het schooljaar zijn afgenomen. Op grond van het bovenstaande is het zeer aannemelijk te veronderstellen, dat de IJkingsgroep op de vijf ontwikkelde tests lager gescoord zou hebben dan de Onderzoeksgroep. Daarmee is tevens waarschijnlijk, dat de prestaties van de Onderzoeksgroep hoger liggen dan die van alle Nederlandse scholen (leerjaar 4) voor gewoon lager onderwijs, aangenomen dat de prestaties sedert 1970-1971 niet zijn gestegen.

We zullen nu de prestaties presenteren van de leerlingen van de acht onderzoeksscholen op de vijf ontwikkelde tests. Tabel 5.2. geeft een overzicht van de belangrijkste gegevens van de A-vormen van de tests. De figuren 5.1. tot en met 5.6. tonen de frequentieverdelingen van de scores op de A-vormen van de tests. De gegevens hebben betrekking op de voormeting (november/december 1978). Opgemerkt moet worden dat de Free Response Cloze Test op twee manieren is nagekeken nl. volgens de zogenaamde exacte woord scoring (FRCT1) en volgens de zogenaamde semantisch juiste scoring (FRCT12).

De frequentieverdelingen van de scores werden met behulp van een chi-kwadraat getoetst op normaliteit. Alle verdelingen bleken op 5% niveau significant van de normaalverdeling af te wijken met uitzondering van de Free Response Cloze Test. Zowel de FRCT1 ($.10 < p < .15$) als de FRCT12 ($.05 < p < .10$) wijken niet significant af. De frequentieverdelingen van de scores op de verschillende tests zijn bevredigend van vorm. De enige test die een duidelijk scheve verdeling vertoont is de Redeneertest. De verde-

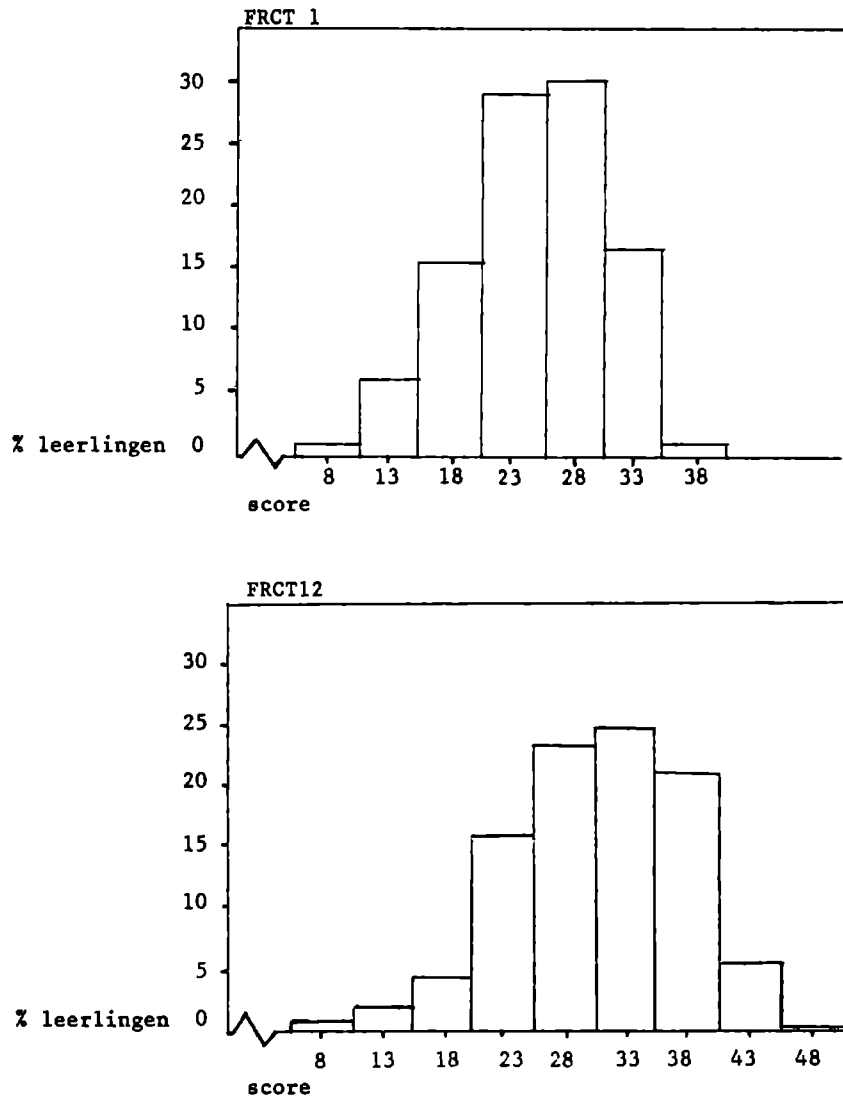
ling van de totaalscores op deze test is sterk positief scheef (rechts asymmetrisch), hetgeen erop wijst dat er veel moeilijke items zijn opgenomen. De scheefheid wordt vooral veroorzaakt door de zogenaamde "onzekere" items. Dit blijkt uit de frequentieverdeling van de subtestscores (fig. 5.6.).

Tabel 5.2. De prestaties van de leerlingen op de ontwikkelde tests.
(vorm A; voormeting: alle scholen). N = ca. 500.

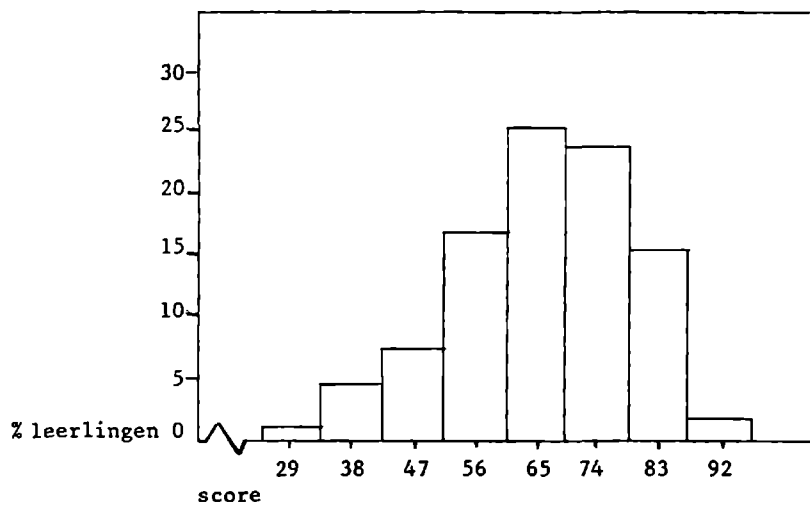
	FRCT1	FRCT12	MCCT	HGD
aantal lln.	520	520	505	497
aantal items	50	50	96	29
gemid. score	24.8	30.5	66.1	14.6
gemid. p.	.50	.61	.69	.50
mediaan	25.3	31.0	67.7	14.1
stand. dev.	5.8	7.1	13.3	7.0

	BED	RED	ONZEKER	ZEKER
aantal lln.	487	496	496	496
aantal items	24	40	16	24
gemid. score	13.6	18.5	6.5	12.0
gemid. p.	.57	.46	.41	.50
mediaan	13.8	16.7	6.1	11.4
stand. dev.	5.2	6.9	4.0	4.9

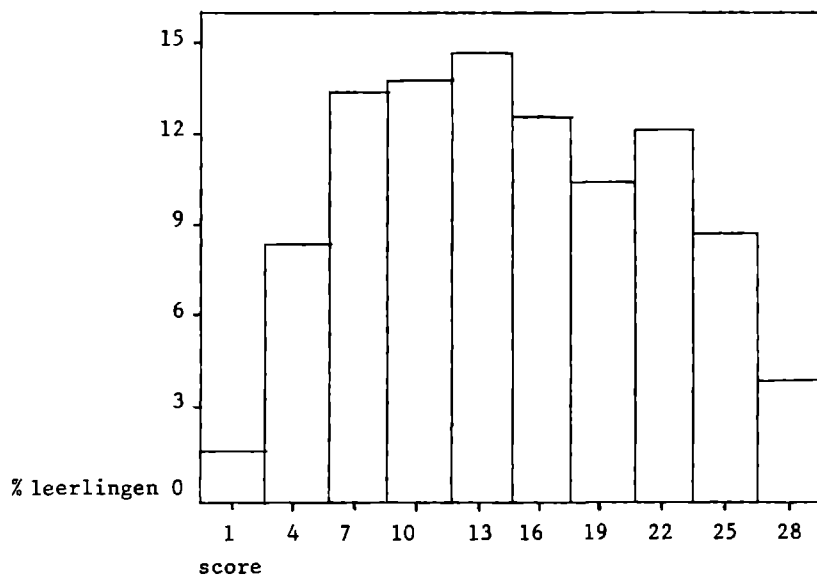
Figuur 5.1. Frequentieverdeling van de Free Response Cloze Test.
(vorm A; voormeting: alle scholen).



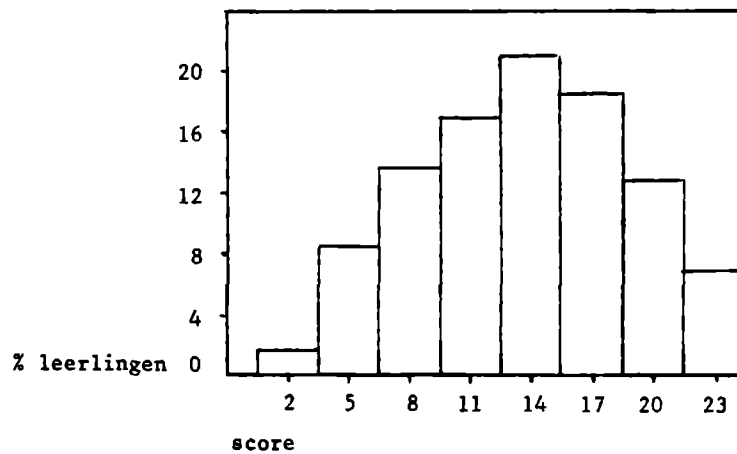
Figuur 5.2. Frequentieverdeling van de Multiple Choice Cloze Test.
(vorm A; voormeting: alle scholen).



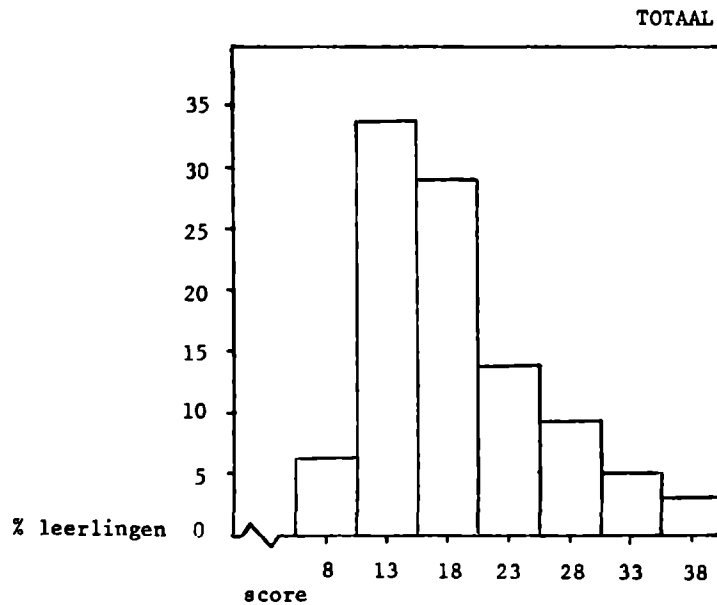
Figuur 5.3. Frequentieverdeling van de Hoofdgedachtetest.
(vorm A; voormeting: alle scholen).



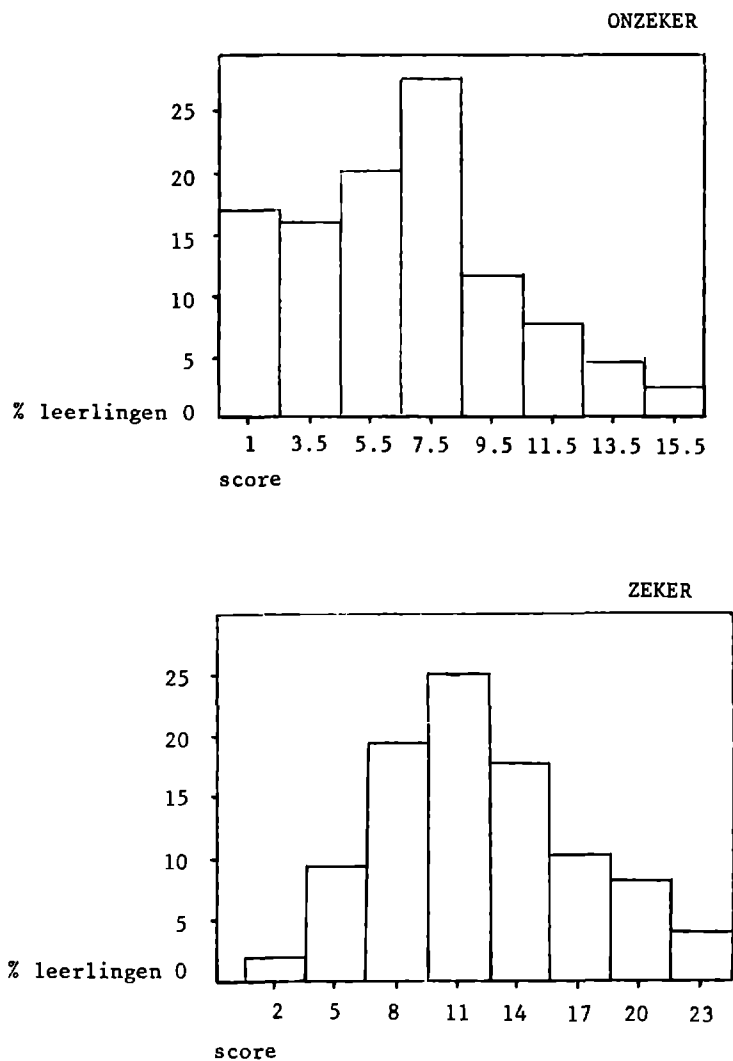
Figuur 5.4. Frequentieverdeling van de Bedoelingtest.
(vorm A; voormeting: alle scholen).



Figuur 5.5. Frequentieverdeling van de Redeneertest.
(vorm A; voormeting: alle scholen).



Figuur 5.6. Frequentieverdeling van de Redeneertest: Onzeker en Zeker.
(vorm A; voormeting: alle scholen).



Verschillen in prestaties tussen sociale milieus

Zoals reeds eerder is opgemerkt, werden de leerlingen ingedeeld naar sociaal milieu op basis van het beroep van de vader en volgens het classificatiesysteem dat ontwikkeld is door het Instituut voor Toegepaste Sociologie te Nijmegen. Tabel 5.3. geeft voor elke test een overzicht van de gemiddelden en standaarddeviaties per sociaal milieu (SM).

Voor elke test werd een variantie-analyse uitgevoerd om na te gaan of de gemiddelde testcores van de sociale milieus significant van elkaar verschillen. Uit deze analyse bleek dat bij alle tests de gemiddelde scores van de sociale milieus zeer significant van elkaar verschilden. Naast deze significantie-toetsing werd voor elke test ook de omvang van de verschillen tussen de sociale milieus berekend. Hierbij werd gebruik gemaakt van het kwadraat van de correlatie-ratio, η^2 . Dit kwadraat geeft aan hoeveel percent van de variantie verklaard wordt door de verschillen in sociale milieus. De kwadraten van de correlatie-ratio variëren per test van maximaal 11% voor de Hoofdgedachtetest tot minimaal 7% voor de Multiple Choice Cloze Test.

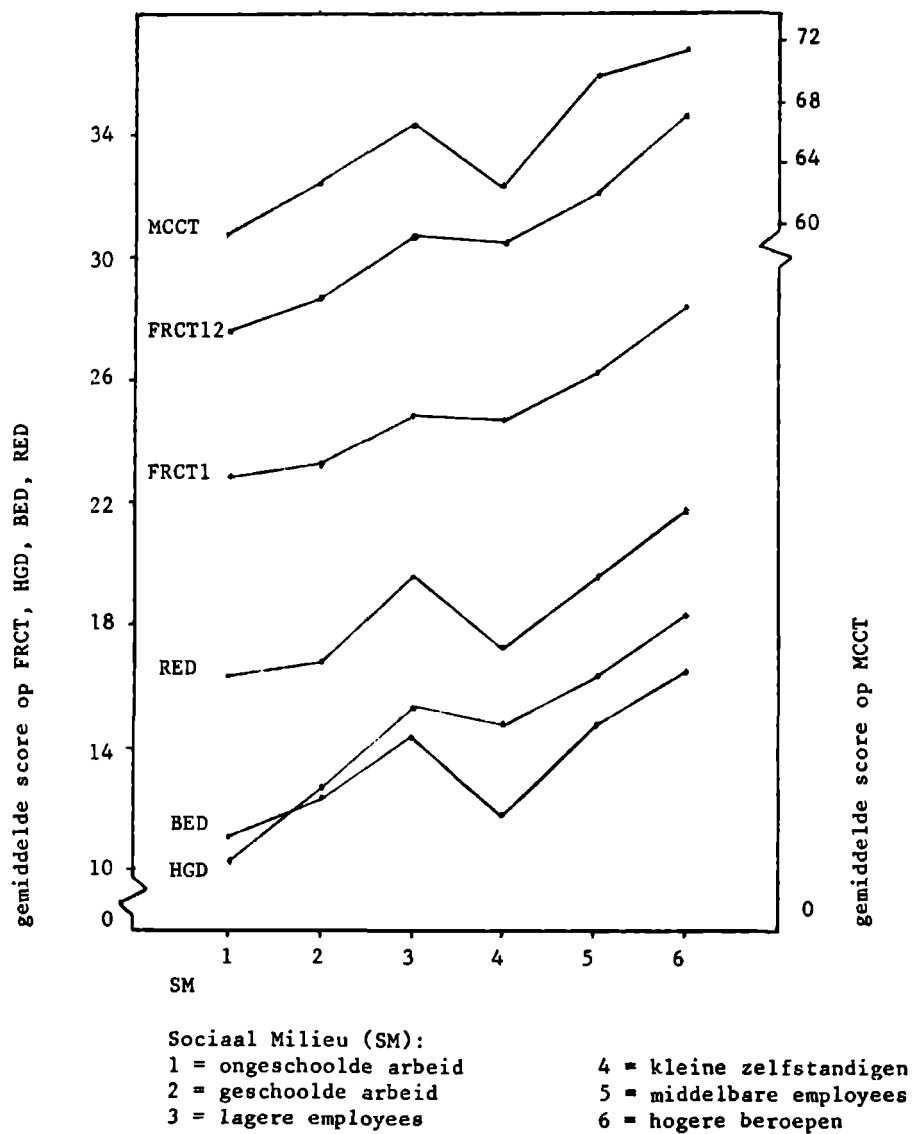
In figuur 5.7. zijn voor alle tests de gemiddelde scores per sociaal milieu grafisch weergegeven. De figuur vertoont een regelmatig patroon. Stijging van het sociaal milieu gaat in het algemeen gepaard met een stijging van de gemiddelde testprestatie. De rangordening van de milieugroepen komt echter niet geheel overeen met de rangorde in de testgemiddelden. Met name de leerlingen van groep 3 (lagere employees) blijken gemiddeld telkens een betere prestatie te leveren dan die van groep 4 (kleine zelfstandigen). Deze bevinding werd reeds eerder geconstateerd door Brus en Voeten (1973), Brus en Van Bergen (1973) en Stijnen (1975).

Tabel 5.3. Gemiddelden en standaarddeviaties per sociaal milieu.
(vorm A; voormeting: alle scholen). N = ca. 475.

SM	FRCT12			FRCT1			MCCT		
	\bar{x}	s	n	\bar{x}	s	n	\bar{x}	s	n
1	27.7	6.8	40	22.9	5.4	40	61.7	11.0	37
2	28.7	7.5	154	23.3	6.2	154	63.2	13.6	145
3	30.8	7.0	101	24.9	5.9	101	66.8	13.4	100
4	30.5	5.9	42	24.7	4.4	42	62.6	12.8	42
5	32.0	6.2	81	26.1	5.2	81	69.9	12.7	80
6	34.8	6.0	75	28.3	5.0	75	71.3	12.7	77
Totaal	30.7	7.1	493	25.0	5.8	493	66.2	13.4	481
F=10.41 p<.0001			F = 9.90 p<.0001			F = 6.71 p<.0001			
bij 5 en 487 vrijheidsgraden.			bij 5 en 487 vrijheidsgraden.			bij 5 en 475 vrijheidsgraden.			

SM	HGD			BED			RED		
	\bar{x}	s	n	\bar{x}	s	n	\bar{x}	s	n
1	10.3	6.2	38	11.1	5.5	36	16.3	5.4	38
2	12.8	6.4	140	12.4	4.9	143	16.9	5.8	145
3	15.2	6.9	101	14.5	5.5	94	19.7	7.5	97
4	14.8	6.5	38	11.9	4.1	39	17.3	6.0	41
5	16.2	6.8	78	14.8	4.7	73	19.5	7.0	76
6	18.3	6.7	78	16.4	4.9	77	21.9	8.0	74
Totaal	14.7	7.0	473	13.7	5.2	462	18.7	7.0	471
F=11.22 p<.0001			F=10.45 p<.0001			F=7.27 p<.0001			
bij 5 en 467 vrijheidsgraden.			bij 5 en 456 vrijheidsgraden.			bij 5 en 465 vrijheidsgraden.			

Figuur 5.7. Gemiddelde testscores per sociaal milieu.
(vorm A; voormeting: alle scholen). N = ca. 475.



Verschillen in prestaties tussen scholen

In tabel 5.4. worden voor elke test de gemiddelden en standaarddeviaties van de acht bij het onderzoek betrokken scholen weergegeven. Voor elke test werd een variantie-analyse uitgevoerd om na te gaan of de schoolgemiddelden significant van elkaar verschillen.

Uit tabel 5.4. blijkt dat bij alle tests de schoolgemiddelden significant van elkaar verschillen ($p < .05$). De significantieniveaus (p-waarden) geven aan dat de verschillen tussen de scholen voor de ene test groter zijn dan voor de andere test. Relatief grote verschillen tussen scholen treden op bij de Multiple Choice Cloze Test, de Hoofdgedachtetest en de Bedoelingtest; relatief kleine verschillen tussen scholen bestaan er bij de Free Response Cloze Test en de Redeneertest. Naast deze significantietoetsing werd ook nu voor elke test de omvang van de verschillen tussen de scholen berekend met behulp van het kwadraat van de correlatie-ratio, η^2 . De kwadraten van de correlatie-ratio variëren per test van maximaal 7% voor de Hoofdgedachtetest en de Multiple Choice Cloze Test tot minimaal 3% voor de Redeneertest en de Free Response Cloze Test 1,12.

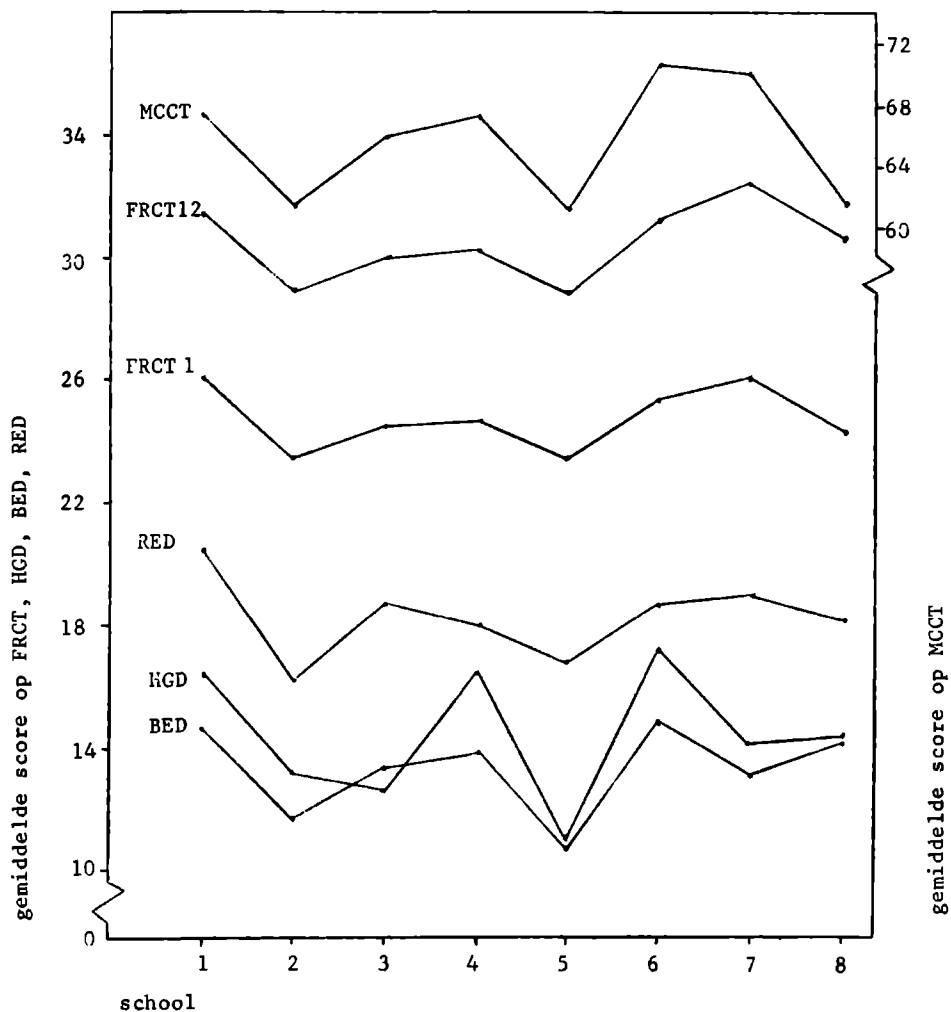
De schoolgemiddelden uit tabel 5.4. worden grafisch weergegeven in figuur 5.8. De curves in deze figuur vertonen een vrij regelmatig beeld. Over de diverse tests hebben vrijwel steeds dezelfde scholen een lage gemiddelde score (de scholen 2 en 5), terwijl andere scholen vrijwel steeds een hoge gemiddelde score hebben (de scholen 1, 6, 7).

Tabel 5.4. Gemiddelden en standaarddeviaties per school.
(vorm A; voormeting: alle scholen). N = ca. 500.

school	FRCT12			FRCT1			MCCT		
	\bar{x}	s	n	\bar{x}	s	n	\bar{x}	s	n
1	31.5	7.0	67	26.1	5.4	67	67.6	13.3	65
2	28.9	7.5	55	23.3	6.1	55	61.7	15.0	55
3	30.0	7.5	74	24.4	6.0	74	66.1	10.6	72
4	30.3	7.7	47	24.5	6.6	47	67.4	14.7	47
5	28.4	6.6	50	23.5	5.6	50	61.0	11.9	44
6	31.1	8.2	61	25.4	6.7	61	70.5	13.7	60
7	32.3	6.0	86	26.1	5.1	86	70.2	10.1	85
8	30.3	6.3	79	24.4	5.3	79	61.9	14.3	77
Totaal	30.5	7.1	519	24.8	5.8	519	66.1	13.3	505
F=2.22 p<.05			F=2.16 p<.05			F=5.44 p<.0001			
bij 7 en 511 vrijheidsgraden.			bij 7 en 511 vrijheidsgraden.			bij 7 en 497 vrijheidsgraden.			

school	HGD			BED			RED		
	\bar{x}	s	n	\bar{x}	s	n	\bar{x}	s	n
1	16.4	6.3	64	14.8	4.9	59	20.4	8.0	65
2	13.2	6.1	53	11.7	4.9	47	16.3	5.5	54
3	12.8	6.9	71	13.3	5.1	73	18.9	6.3	69
4	16.6	7.1	47	14.0	5.6	41	18.0	7.0	45
5	11.1	6.5	45	10.9	4.9	47	16.9	6.1	47
6	17.4	7.5	61	15.2	5.5	60	18.9	7.8	61
7	14.3	7.2	78	13.6	4.9	80	19.1	6.6	80
8	14.6	6.1	78	14.4	5.1	80	18.4	7.1	75
Totaal	14.6	7.0	497	13.6	5.2	487	18.5	6.9	496
F=5.59 p<.0001			F=4.52 p<.0005			F=2.09 p<.05			
bij 7 en 489 vrijheidsgraden.			bij 7 en 479 vrijheidsgraden.			bij 7 en 488 vrijheidsgraden			

Figuur 5.8. Gemiddelde testscore per school.
(vorm A; voormeting: alle scholen). N = ca. 500.



Tot slot moet worden opgemerkt dat de gegevens zoals boven vermeld alleen een indicatie geven omtrent de grootte van de verschillen tussen de gemiddelde testcores van de onderscheiden sociale milieus en scholen. Op grond van deze gegevens kunnen geen uitspraken worden gedaan omtrent het relatieve belang van de onderling afhankelijke factoren "sociaal milieu"

en "school". Omdat in dit onderzoek slechts enkele gegevens zijn verzameld omtrent het sociaal milieu en de scholen wordt hier geen poging gedaan de invloeden van beide factoren op de testprestaties te analyseren, rekening houdend met de onderlinge afhankelijkheid van deze factoren.

5.1.3. De betrouwbaarheid van de ontwikkelde tests

Opzet

Zoals uit de voorgaande hoofdstukken blijkt, is de vaststelling van de betrouwbaarheid van de ontwikkelde tests een belangrijk doel van het descriptief gedeelte van het onderzoek. Dat is de reden dat in deze paragraaf vrij uitvoerig wordt ingegaan op de betrouwbaarheid van de tests.

De betrouwbaarheid van een test heeft betrekking op de mate waarin men op de resultaten van een meting staat kan maken. Hoe groter de invloed van toeval op de meetresultaten, hoe lager de betrouwbaarheid van de test. Een bespreking van de betrouwbaarheid van een test wordt gecompliceerd door het feit, dat er niet één maat bestaat die eens en voor al aangeeft hoe betrouwbaar een test is. In de eerste plaats zijn er verschillende betrouwbaarheidsmaten die elk een enigszins ander aspect van de betrouwbaarheid meten. In de tweede plaats zijn de waarden die deze maten aannemen afhankelijk van de onderzochte populatie leerlingen en het tijdstip waarop gemeten wordt.

Drenth (1972) onderscheidt vier hoofdmethoden voor de bepaling van de betrouwbaarheid. Deze leveren elk een eigen betrouwbaarheidscoëfficiënt op. Voor de eerste twee methoden is een herhaald testonderzoek nodig, bij de laatste twee methoden kan worden volstaan met één afname. Hieronder volgt een opsomming van de methoden en de daarbij horende betrouwbaarheidscoëfficiënten.

- a. De test-hertestmethode, waarbij twee maal dezelfde test wordt afgenomen. De correlatie tussen de beide afnames levert een coëfficiënt van stabiliteit.
- b. De parallelvorm-methode, waarbij de tweede meting bestaat uit een parallelvorm van de eerste meting. De correlatie tussen de parallelvormen levert een coëfficiënt van a-specificiteit.
- c. De splitsingsmethode, waarbij de verzameling testitems in twee gelijke helften gesplitst wordt. De correlatie tussen de helften levert een co-

efficiënt van interne consistentie.

- d. De methode van item-analyse, waarbij de test in evenveel delen verdeeld wordt als er items zijn. De intercorrelaties van de items leveren een coëfficiënt van item-consistentie, die een variant is van de coëfficiënt van interne consistentie.

De opzet van het onderzoek staat toe dat alle vier methoden gebruikt worden.

- In de voormeting en de retentiemeting zijn de A-vormen van de tests afgenomen, zodat correlaties tussen voormeting en retentiemeting opgevat kunnen worden als stabiliteitscoëfficiënten.
- De nameting bestaat uit de parallelle vormen van de voormeting i.c. de B-vormen en de retentiemeting bevat de parallelle vormen van de nameting i.c. de A-vormen. Correlaties tussen voormeting (A-vormen) en nameting (B-vormen) en tussen nameting (B-vormen) en retentiemeting (A-vormen) kunnen dus opgevat worden als coëfficiënten van a-specificiteit. Gezien de vrij lange tijd tussen de metingen (enkele maanden) zit er echter tevens een stabiliteitsaspect aan deze correlaties.
- De drie taakgerichte tests bestaan uit twee delen nl. A1 en A2, respectievelijk B1 en B2. De correlatie tussen de delen kan opgevat worden als een coëfficiënt van interne consistentie. Omdat deze delen zo gelijk mogelijk zijn samengesteld, kan de correlatie tussen de delen ook worden opgevat als een coëfficiënt van a-specificiteit. Daar de betreffende coëfficiënt gebaseerd is op halve tests, wordt de correctieformule van Spearman-Brown gebruikt om een schatting te krijgen van de betrouwbaarheid van de hele test. De FRCT en de MCCT bestaan niet uit twee delen, zodat voor deze tests geen correlaties tussen delen gegeven worden.
- Binnen elk meetmoment afzonderlijk (voormeting, nameting, retentiemeting) kunnen coëfficiënten van item-consistentie berekend worden. Voor de berekening van de item-consistentie wordt de veel gebruikte formule nr. 20 van Kuder en Richardson, de KR-20, gehanteerd.

Een probleem bij de berekening van de item-consistentie vormen de cloze tests, omdat de items van deze tests experimenteel niet onafhankelijk zijn. De kans dat een leerling een item goed maakt, zal er mede van afhankelijk zijn of hij andere, met name aangrenzende, items goed beantwoord heeft. Een betrouwbaarheidsschatting, gebaseerd op inter-item-correlaties zoals de KR-20 zal in zo'n geval een geflatteerd beeld te zien geven. Voor de MCCT

is dit probleem op te lossen door de interne consistentie te berekenen over de subtests en niet over de items. De MCCT bestaat uit 8 subtests (teksten). Deze subtests zijn onderling wél experimenteel onafhankelijk. Coëfficiënt alpha geeft een schatting van de interne consistentie, gebaseerd op de correlaties tussen subtests. Voor de FRCT is geen goed alternatief beschikbaar. Een splitsingsmethode zou ook geen oplossing bieden. Daarom is voor de FRCT uiteindelijk toch gekozen voor de KR-20. De gevonden betrouwbaarheid kan vanwege de afhankelijkheid van de items enigszins geïflatteerd zijn.

De correlatie tussen de testdelen kan vergeleken worden met de KR-20. De KR-20 is een schatting van het gemiddelde van alle mogelijke split-halves. Hoewel de delen zo zijn samengesteld dat ze qua moeilijkheidsgraad, spreiding en inhoud zo veel mogelijk aan elkaar gelijk zijn, kan de correlatie tussen de delen lager zijn dan de KR-20. De delen zijn op dezelfde dag met enige tussentijd afgenomen. Veranderingen in die tussentijd, vermoeidheid en tegenzin bij het tweede deel, kunnen de correlatie tussen de delen negatief beïnvloeden.

Zoals opgemerkt is de berekende betrouwbaarheid niet alleen afhankelijk van de gebruikte methode maar ook van het tijdstip waarop gemeten wordt. De momenten waarop gemeten is, zijn november/december (voormeting), maart (nameting) en mei/juni (retentiemeting). De betrouwbaarheid van de tests kan op deze momenten verschillen, met name als gevolg van de omstandigheid dat de leerlingen veranderen in datgene wat de tests meten.

Ook de onderzochte populatie is van invloed op de betrouwbaarheidscoëfficiënten van de tests. Het is niet onwaarschijnlijk dat enigszins andere coëfficiënten worden gevonden wanneer andere scholen voor het onderzoek geselecteerd zouden zijn. Bovendien is het in dit verband van belang, dat een deel van de leerlingen een experimenteel programma heeft gevolgd dat er expliciet op gericht is verbetering te bewerkstelligen in datgene wat de Hoofdgedachtetest, de Bedoelingtest en de Redeneertest meten. Om deze laatste reden lijkt het nuttig de geselecteerde scholen te onderscheiden in Experimentele Scholen die een experimenteel programma gehad hebben en Controlescholen die geen experimenteel programma gehad hebben. De leerlingen van de Experimentele Scholen vormen een heterogene groep; 1/4 deel van de leerlingen heeft het programma Hoofdgedachte gevolgd, 1/4 deel het programma Bedoeling, 1/4 deel het programma Redeneren en 1/4 deel het Controlepro-

gramma. Het is mogelijk om de leerlingen van de Experimentele Scholen verder op te splitsen naar de aard van het gevolgde programma (Hoofdgedachtegroep, Redeneergroep, etc.). Daar het aantal leerlingen per groep dan erg klein zou worden voor een betrouwbaarheidsonderzoek, is van een opsplitsing van de leerlingen van de Experimentele Scholen afgezien. De vraag op welke populatie de betrouwbaarheidsmaten betrekking hebben, wordt enigszins gecompliceerd door de omstandigheid dat in de loop van het onderzoek twee scholen uitvielen of geen bruikbare gegevens opleverden. Van de experimentele school 2 zijn alleen de gegevens van de voormeting bruikbaar; de gegevens van de nameting en de retentiemeting van deze school zullen buiten beschouwing worden gelaten. Van de controleschool 5 zijn wel gegevens beschikbaar van de voormeting en de nameting, maar niet van de retentiemeting.

Besloten is betrouwbaarheidsmaten te berekenen van vier groepen scholen nl.

- De Totale Groep maximaal. Voor elk meetmoment worden berekeningen uitgevoerd over alle scholen waarvan gegevens beschikbaar zijn. Bij de voormeting bestaat deze groep uit alle scholen (1 tot en met 8), bij de nameting uit alle scholen minus school 2 en bij de retentiemeting uit alle scholen minus de scholen 2 en 5.
- De Totale Groep constant. Om vergelijkingen tussen meetmomenten beter mogelijk te maken, is ook een totale groep gevormd die over de meetmomenten constant blijft. De "Totale Groep constant" bestaat uit alle scholen minus de scholen 2 en 5.
- De Experimentele Scholen. In principe zou hier analoog aan de totale groep onderscheid gemaakt kunnen worden tussen "experimentele scholen maximaal" en "experimentele scholen constant". De voorkeur werd gegeven aan een constante verzameling experimentele scholen om vergelijkingen over meetmomenten beter mogelijk te maken. De Experimentele Scholen bestaan derhalve voor alle meetmomenten uit de scholen 1, 3 en 4.
- De Controlescholen. Evenals bij de Experimentele Scholen wordt ook bij de Controlescholen afgezien van een groep "controlescholen maximaal". De controlescholen bestaan voor alle meetmomenten uit de scholen 6, 7 en 8.

Een aantal gevonden betrouwbaarheidscoëfficiënten kan vergeleken worden met coëfficiënten berekend over andere populaties nl. de leerlingen die deelgenomen hebben aan de afnames ten behoeve van de constructie van de pa-

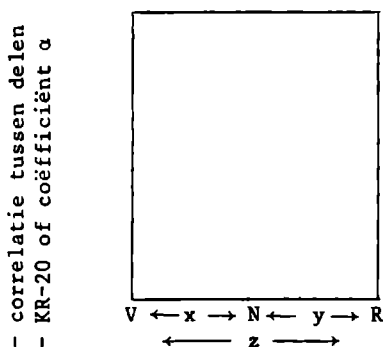
rallelvormen. Bij deze afnames verschilden de populaties per test. Daar de tests toen als één geheel zijn afgenomen, zijn hier géén test-hertest gegevens beschikbaar. De correlaties tussen de parallelvormen zijn geflatteerd, omdat de vormen achteraf op basis van de resultaten van de afnames werden geconstrueerd (vgl. par. 4.4.).

Resultaten

Het voorgaande heeft geleid tot het verzamelen van de volgende gegevens betreffende de betrouwbaarheid van de tests. Bij elke test worden vier groepen scholen onderscheiden nl. de Totale Groep maximaal, de Totale Croep constant, de Experimentele Scholen en de Controlescholen. Voor elke groep wordt per meetmoment het aantal leerlingen vermeld, de gemiddelde p-waarde, de standaarddeviatie, de KR-20 (respectievelijk alpha-coëfficiënt), de correlatie tussen de testdelen (indien van toepassing) en de correlatie tussen de testdelen, gecorrigeerd volgens Spearman-Brown. Ook worden voor elke groep de correlaties tussen de meetmomenten gepresenteerd. Tenslotte worden resultaten gegeven van de afnames ten behoeve van de constructie van parallelvormen. Deze gegevens zijn te vinden in Bijlage 5.1.

De belangrijkste resultaten uit Bijlage 5.1. zijn grafisch weergegeven in figuur 5.9. Deze figuur bevat voor elke test het verloop over de meetmomenten van de KR-20 (respectievelijk alpha-coëfficiënt) en van de correlatie tussen de testdelen gecorrigeerd volgens Spearman-Brown (indien van toepassing). Deze curves worden gepresenteerd voor drie groepen scholen nl. de Totale Groep constant, de Experimentele Scholen en de Controlescholen. Tevens worden bij de grafieken de correlaties tussen de meetmomenten weergegeven.

Figuur 5.9. Betrouwbaarheidsmaten voor de ontwikkelde tests.



INDEX

———— = KR-20 (voor MCCT: coëff. α)

----- = Correlatie tussen testdelen

V = Voormeting

N = Nameting

R = Retentiemeting

V \leftrightarrow x \leftrightarrow N = Correlatie tussen V en N

N \leftrightarrow y \leftrightarrow R = Correlatie tussen N en R

V \leftrightarrow z \leftrightarrow R = Correlatie tussen V en R

Totale Groep constant = school 1, 3, 4, 6, 7, 8
(N = ca. 400)

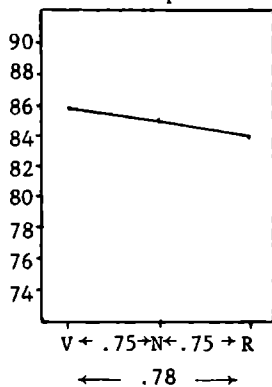
Exp. scholen = school 1, 3, 4 (N=ca. 185)

Controlescholen = school 6, 7, 8
(N=ca. 220).

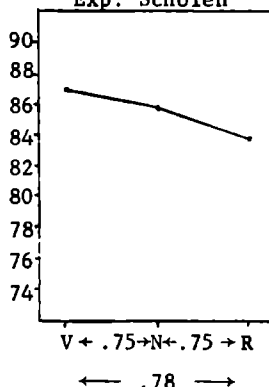
Figuur 5.9. Vervolg.

FRCT 12

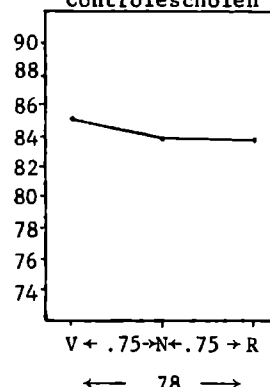
Totale Groep constant



Exp. Scholen

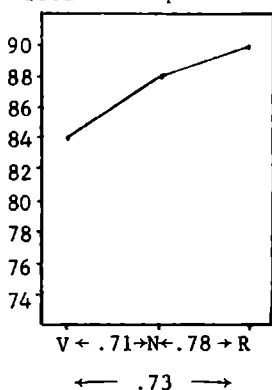


Controlescholen

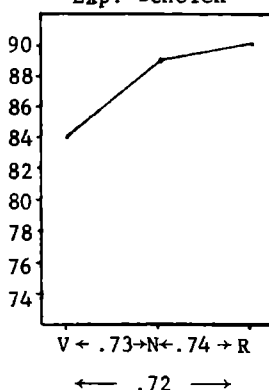


MCCT

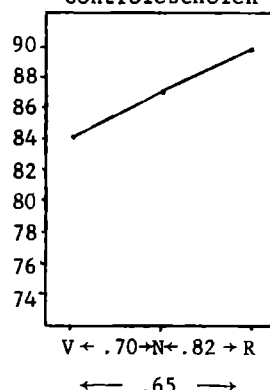
Totale Groep constant



Exp. Scholen

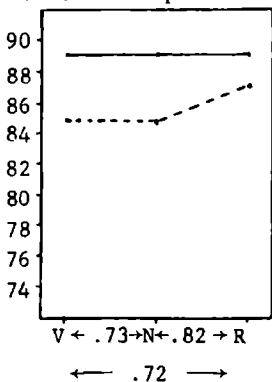


Controlescholen

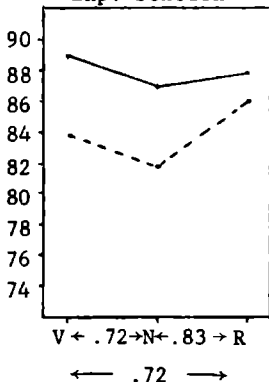


HOOFDGEDACHTETEST

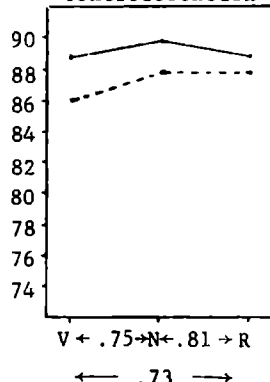
Totale Groep constant



Exp. Scholen



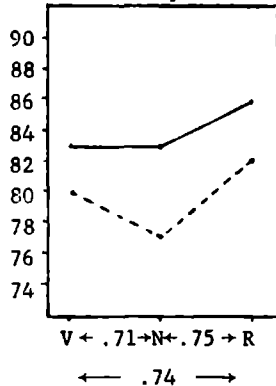
Controlescholen



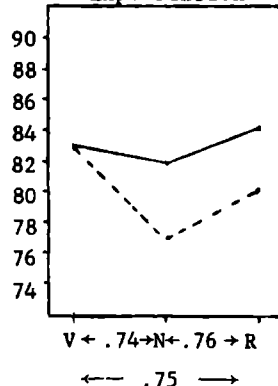
Figuur 5.9. Vervolg.

BEDOELINGTEST

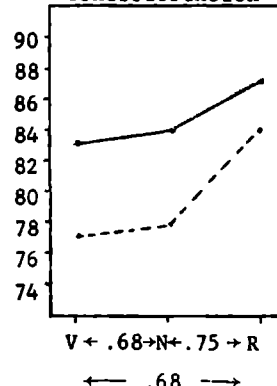
Totale Groep constant



Exp. Scholen

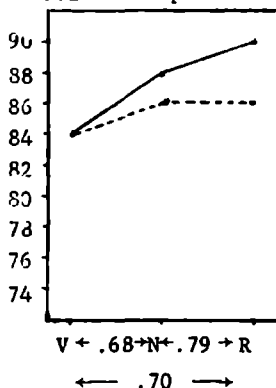


Controlescholen

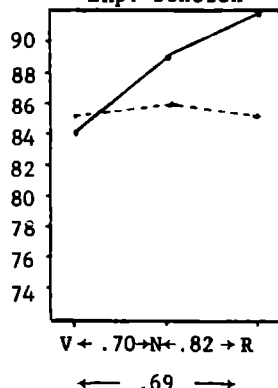


REDENEERTEST: TOTAAL

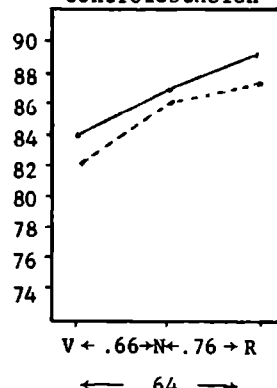
Totale Groep constant



Exp. Scholen

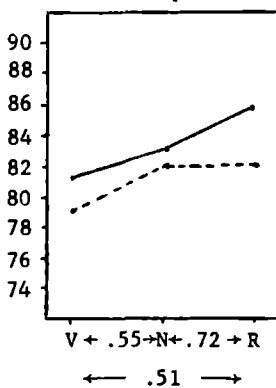


Controlescholen

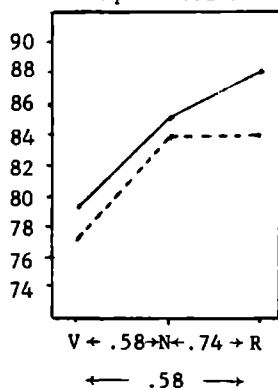


REDENEERTEST: ZEKER

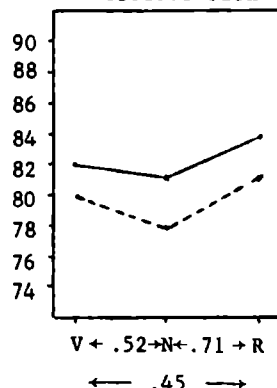
Totale Groep constant



Exp. Scholen

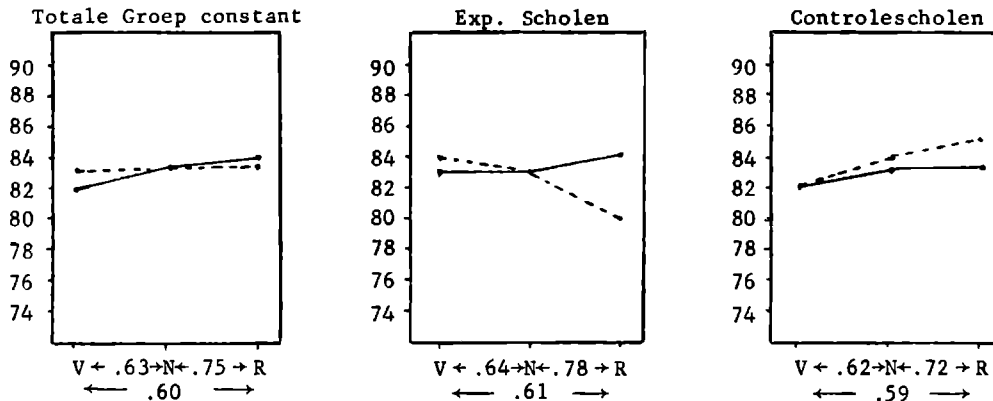


Controlescholen



Figuur 5.9. Vervolg.

REDENEERTEST: ONZEKER



Free Response Cloze Test 12

De KR-20 van deze test is voor alle scholen inclusief 2 en 5 bij de voormeting .86. De KR-20 blijft over de meetmomenten heen vrijwel gelijk met een lichte tendens tot dalen. Deze lichte daling wordt bij de retentiemeting waarschijnlijk veroorzaakt door een afname van de spreiding.

De correlaties tussen de meetmomenten vertonen een opmerkelijk patroon nl. de correlatie tussen voormeting en retentiemeting is het hoogst, ondanks het feit dat tussen deze twee metingen de meeste tijd ligt. Mogelijk ligt een verklaring hiervoor in de omstandigheid, dat bij de voormeting en de retentiemeting dezelfde testvormen werden afgenomen, terwijl de beide andere correlaties betrekking hebben op verschillende testvormen. Dit zou betekenen dat de prestaties van de leerlingen enigszins afhankelijk zijn van de inhoud van de FRCT (vorm A of vorm B). Absoluut gezien zijn de correlaties tussen de meetmomenten vrij hoog. De prestaties van de leerlingen zijn dus over de tijd gezien vrij stabiel, ook als de FRCT een andere inhoud heeft (vorm A versus vorm B).

Multiple Choice Cloze Test

De alpha-coëfficiënt van deze test is voor alle scholen inclusief 2 en 5 bij de voormeting .85. De coëfficiënt stijgt over de meetmomenten naar

.90 bij de retentiemeting. Dit geldt voor alle vier groepen van scholen.

Uit de absolute hoogte van de correlaties tussen de meetmomenten valt af te lezen dat de MCCT vrij stabiel is. De correlaties tussen de meetmomenten vertonen een kenmerk dat ook bij andere tests optreedt nl. de correlatie nameting-retentiemeting is hoger dan de correlaties voormeting-nameting en voormeting-retentiemeting. Wat betreft de MCCT is dit althans het geval bij de Totale Groep en bij de Controlescholen; bij de Experimentele Scholen zijn de verschillen te verwaarlozen. De vraag is hoe het komt dat de correlatie nameting-retentiemeting bij de MCCT hoger is.

De verschillen tussen de correlaties kunnen niet verklaard worden uit verschillen in interne consistentie van de test op de drie meetmomenten. De verschillen tussen de correlaties kunnen ook niet verklaard worden door de testvormen waar tussen gecorreleerd wordt, want bij de correlatie voormeting-nameting en de correlatie nameting-retentiemeting wordt in beide gevallen de correlatie berekend tussen vorm A en vorm B. Een mogelijke verklaring voor de hogere correlatie tussen nameting-retentiemeting en de lagere correlatie tussen voormeting-nameting is de toename van de spreiding tussen de beide eerstgenoemde meetmomenten. Deze verklaring lijkt juist, maar is niet geheel toereikend wanneer ook de correlaties tussen de meetmomenten bij de andere tests in de beschouwing worden betrokken. Zo kan bijvoorbeeld de hogere correlatie tussen nameting-retentiemeting bij de Controlescholen op de Hoofdgedachtetest, de Bedoelingtest en de Redeneertest niet of in onvoldoende mate worden verklaard door een toename in spreiding. Een mogelijke aanvullende verklaring voor met name de lagere correlatie tussen voormeting-nameting en voormeting-retentiemeting is dat tijdens de voormeting een aantal leerlingen prestaties heeft geleverd die beneden hun eigenlijke niveau liggen. In par. 5.1.4. en 5.1.5. komt deze problematiek opnieuw aan de orde.

Hoofdgedachtetest

De KR-20 is voor deze test voor alle scholen inclusief 2 en 5 bij de voormeting .89. De KR-20 blijft voor alle groepen scholen vrijwel constant over de meetmomenten.

De correlatie tussen de testdelen is bij de voormeting voor alle scholen inclusief 2 en 5 en gecorrigeerd volgens Spearman-Brown .86. Deze correla-

tie vertoont bijna eenzelfde verloop over de meetmomenten als de KR-20. Verschillen tussen de Experimentele Scholen en de Controlescholen hangen samen met verschillen in spreiding. Zo hangt de relatief lage correlatie bij de nameting van de Experimentele Scholen samen met een afname van de spreiding.

De correlaties tussen de meetmomenten vertonen eenzelfde patroon als bij de MCCT nl. de correlatie nameting-retentiemeting is hoger dan de correlaties voormeting-nameting en voormeting-retentiemeting. Dit verschijnsel treedt hier ook op bij de Experimentele Scholen. Uit de absolute hoogte van de correlaties blijkt dat de Hoofdgedachtetest vrij stabiel is.

Bedoelingtest

De KR-20 voor deze test is voor alle scholen inclusief 2 en 5 bij de voormeting .83. Bij de nameting treedt voor de Experimentele Scholen een lichte daling op in de KR-20 en bij de Controlescholen een lichte stijging. De KR-20 bereikt de hoogste waarde bij de retentiemeting van de Controlescholen nl. .87.

De correlatie tussen de testdelen voor alle scholen inclusief 2 en 5 is bij de voormeting, gecorrigeerd volgens Spearman-Brown, .81. Het verloop van de correlaties over de meetmomenten verschilt tussen de Experimentele Scholen en de Controlescholen. Dit verschil is gerelateerd aan verschillen in spreiding. Het verloop van de correlatie over de meetmomenten lijkt op dat van de KR-20; de correlatie is doorgaans iets lager dan de KR-20.

De correlaties tussen de meetmomenten zijn hoger bij de nameting-retentiemeting dan bij de voormeting-nameting en de voormeting-retentiemeting. Dit geldt met name voor de Controlescholen. Evenals bij de MCCT zijn de verschillen in correlaties bij de Experimentele Scholen te verwaarlozen. Uit de absolute hoogte van de correlaties blijkt, dat de Bedoelingtest vrij stabiel is; de stabiliteit is wat lager voor de Controlescholen.

Redeneertest: Totaal

De KR-20 voor de totale test is voor alle scholen inclusief 2 en 5 bij de voormeting .83. De KR-20 vertoont voor alle groepen scholen een stijging over de meetmomenten. Toch kan niet worden aangenomen dat de betrouwbaarheid van de Redeneertest sterk toeneemt over de meetmomenten. De toename van de KR-20 coëfficiënten is waarschijnlijk voor een groot deel een gevolg

van een grotere spreiding in de verschillende groepen bij de nameting en retentiemeting.

De correlatie tussen de testdelen ligt dicht bij de KR-20. Bij de Experimentele Scholen treedt bij de retentiemeting een wat groter verschil op tussen deze correlatie en de KR-20. De oorzaak hiervan is onduidelijk.

Bij de correlatie tussen de meetmomenten is de correlatie nameting-retentiemeting relatief hoog. Dit geldt voor alle vier groepen van scholen. De correlaties voormeting-nameting en voormeting-retentiemeting zijn beduidend lager. De Redeneertest is een vrij stabiele test; de stabiliteit is bij de Controlescholen lager.

Redeneertest: Zeker

De KR-20 is voor dit deel van de test voor alle scholen inclusief 2 en 5 bij de voormeting .80. Het verloop van de KR-20 over de meetmomenten is voor de Experimentele Scholen en de Controlescholen nogal verschillend. Dit verschil kan worden toegeschreven aan verschillen in spreiding in beide groepen. De toename van de spreiding bij de Experimentele Scholen op de nameting en de retentiemeting is waarschijnlijk een effect van het experimentele programma.

De correlatie tussen de testdelen is wat lager dan de KR-20. Het verloop van de correlatie over de meetmomenten lijkt veel op dat van de KR-20. Bij de correlaties tussen de meetmomenten is de correlatie nameting-retentiemeting veel hoger dan de correlatie voormeting-nameting en voormeting-retentiemeting. Vergeleken met de andere tests, en ook met de gehele Redeneertest zijn de correlaties voormeting-nameting en voormeting-retentiemeting laag. De prestaties op de subtest "Zeker" zijn nogal instabiel. In de tijd tussen voormeting en nameting zijn vrij sterke veranderingen opgetreden in de prestaties van de leerlingen ten opzichte van elkaar. Dit geldt niet alleen voor de Experimentele Scholen maar ook voor de Controlescholen.

Redeneertest: Onzeker

De KR-20 is voor dit deel van de test voor alle scholen inclusief 2 en 5 bij de voormeting .83. De KR-20 blijft evenals de spreiding vrijwel constant over de meetmomenten. In tegenstelling tot de "Redeneertest Totaal" en de Redeneertest "Zeker" treedt bij deze subtest geen toename van de

spreiding op, ook niet bij de Experimentele Scholen.

De correlatie tussen de testdelen is soms iets hoger en soms iets lager dan de KR-20; de verschillen met de KR-20 zijn vrij klein. De correlaties tussen de meetmomenten vertonen het bekende beeld van een hogere correlatie nameting-retentiemeting vergeleken met voormeting-nameting en voormeting-retentiemeting. Ook de Redeneertest "Onzeker" is vrij instabiel tussen voormeting en nameting, zij het niet zo sterk als de Redeneertest "Zeker".

Uit de hoogte van de coëfficiënten van stabiliteit, a-specificiteit en vooral van interne consistentie kan worden afgeleid dat de vijf ontwikkelde tests voldoen aan de eisen van betrouwbaarheid. Wat betreft deze coëfficiënten bestaan er verschillen tussen de experimentele en controlescholen. Deze verschillen zijn over het algemeen niet groot en variëren per test en meetmoment.

Scoringswijzen van de FRCT en beoordelaarbetrouwbaarheid

Tot slot van deze paragraaf over de betrouwbaarheid van de tests wordt de beoordelaarbetrouwbaarheid of nakijk-consistentie van de FRCT beschreven. Als enige van de vijf ontwikkelde tests kon de juistheid van de antwoorden op de Free Response Cloze Test niet langs mechanische weg bepaald worden. Alle testformulieren van de FRCT uit zowel de voormeting, de nameting als de retentiemeting zijn door dezelfde persoon nagekeken. Aan elk item kon één van de volgende vier codes worden toegekend:

- een 9 als niets was ingevuld of het ingevulde niet te ontcijferen was;
- een 0 als een foutief antwoord was ingevuld of als meerdere woorden waren ingevuld;
- een 1 als het oorspronkelijke woord op de open plaats was ingevuld;
- een 2 als niet het oorspronkelijke woord, maar wel een in die tekst passend woord was ingevuld.

Uit vooronderzoek was bekend dat vooral bij de laatste code nogal wat inconsistenties voor kunnen komen. Daartoe is voor elk van de 50 items een lijst aangelegd met woorden die de code 2 behoorden te krijgen en woorden die met een 0 gecodeerd zouden moeten worden. Het aantal arbitraire beslissingen van de codeur wordt zo drastisch verminderd.

Als controle op de nakijk-consistentie zijn de testformulieren uit de

voormeting van de leerlingen uit drie klassen ($N = 84$) na ongeveer twee maanden opnieuw door dezelfde codeur nagekeken. De keuze van deze klassen uit de 19 beschikbare was zodanig dat:

- a. de klas die oorspronkelijk als eerste was nagekeken de tweede keer als laatste werd gecodeerd;
- b. de klas die oorspronkelijk als tiende (middelste) was nagekeken, de tweede keer opnieuw als middelste (maar nu als tweede) werd nagekeken;
- c. de klas die oorspronkelijk als laatste was gecodeerd de tweede keer als eerste werd nagekeken.

Met behulp van deze gegevens is de overeenstemming in codering voor de items afzonderlijk en voor alle items gezamenlijk berekend. De percentages overeenstemming voor de items varieerden van 80% tot 100%. Bij het merendeel van de items was het percentage overeenstemming groter dan 95%. De totale nakijk-consistentie bedroeg 97.2%. Voor de analyse van de testgegevens werd nooit van alle vier codes tegelijk gebruik gemaakt. De ene maal werden de items met de code 2 omgecodeerd tot 0; m.a.w. het antwoord wordt fout gerekend. We spreken dan van exacte woordscoring. In dit rapport werd dit tot nu toe aangeduid als FRCT1. De andere maal werden de items met de code 2 omgecodeerd tot 1; m.a.w. het antwoord wordt goed gerekend. We spreken dan van semantisch juiste scoring. We duiden dit aan als FRCT12 (wordt de aanduiding van de scoringswijze achterwege gelaten, dan wordt de semantisch juiste scoring bedoeld).

De eerste keer bleek de nakijk-consistentie 98.4%, de tweede keer 97.9% te zijn. Deze waarden zijn alleszins acceptabel. De aard van de gevonden inconsistenties kan afgelezen worden uit tabel 5.5. Het grootst blijkt de "0 - 2 inconsistentie" te zijn.

Vanaf dit moment zal bij de analyse van de gegevens alleen nog gewerkt worden met de totaal- en itemscore verkregen met behulp van de semantisch juiste scoring. Gebleken is dat de nakijk-consistentie niet noemenswaard verschilt en dat deze wijze van scoring resulteert in hogere gemiddelde waarden en een grotere spreiding. Bij de constructie van de testvormen bleek al dat de semantisch juiste scoringswijze hogere betrouwbaarheidscoëfficiënten en hogere onderlinge correlaties opleverde. Deze manier van nakijken sluit ook het best aan bij de instructie van de test. Bij de behandeling van het voorbeeld wordt namelijk niet gesproken over h~~et~~ goede antwoord, maar van goede antwoorden. Bovendien zal de semantisch juiste sco-

ringswijze de test voor zowel de leerling als de leerkracht meer acceptabel maken.

Tabel 5.5. Overzicht van de nakijk-consistentie van testformulieren van 84 leerlingen na twee maanden (50 items per testformulier).

	code	eerste maal nagekeken				
		0	1	2	9	Totaal
tweede maal nagekeken	0	1452	19	21	1	1493
	1	21	2116	14	0	2151
	2	27	11	444	0	482
	9	1	1	0	72	74
	Totaal	1501	2147	479	73	4200

5.1.4. De validiteit van de ontwikkelde tests

Inleiding

De literatuurstudie die in verband met de selectie en specificatie van de drie inferentiële leestaken is uitgevoerd, had vooral tot doel de validiteit van de drie taakgerichte tests zo veel mogelijk veilig te stellen. Hetzelfde kan gezegd worden van de literatuurstudie die in verband met de cloze tests is uitgevoerd.

Zoals uit hoofdstuk IV blijkt, is er naar gestreefd de drie taakgerichte tests en de twee cloze tests op een systematische wijze te ontwikkelen. Aan één eis kon bij de constructie van deze tests niet geheel worden voldaan, nl. aan de eis die in de testtheorie vaak wordt aangeduid met de term inhoudsvaliditeit (De Zeeuw, 1978). Hoewel de taken van de drie taakgerichte tests zo goed mogelijk werden gespecificeerd en de kenmerken van de teksten voor deze tests en de cloze tests zo duidelijk mogelijk werden omschreven, zijn de teksten van de vijf tests niet op een systematische wijze geselecteerd uit de gangbare leesmethoden. Om aan deze eis te voldoen zou een uitputtende inventarisatie nodig zijn geweest van de teksten die in verband met de te ontwikkelen tests in de verschillende leesmethoden voorkomen. Een dergelijke inventarisatie van het betreffende leesstofuniversum werd niet reëel geacht, gezien de onduidelijke structuur van veel leesmethoden. De teksten van de Hoofdgedachtetest en de Bedoelingtest zijn dan ook niet op een systematische wijze geselecteerd. Hetzelfde geldt voor de teksten van de cloze tests; ook daar is

van een sampling geen sprake geweest. De teksten van de Redeneertest zijn door de constructeurs zelf geschreven. De vraag in hoeverre de inhoud van de verschillende tests bepaalde leesstof i.c. teksten uit de gangbare leesmethoden representeert, kan hier dus niet worden beantwoord.

Een andere en in dit verband meer fundamentele vraag is in hoeverre de tests inderdaad meten wat zij geacht worden te meten. Meten de betreffende tests inderdaad een binnen de leesvaardigheid te onderscheiden aspect? Is dit aspect te beschouwen als begrijpend lezen? Meten de taakgerichte tests een specifiek aspect van begrijpend lezen? Deze en andere vragen kunnen in verband met de begripsvaliditeit van de tests gesteld worden.

In verband met de begripsvaliditeit van de ontwikkelde tests zal de samenhang worden vastgesteld tussen de tests onderling en tussen deze tests en maten voor technisch lezen, begrijpend lezen, woordenschat en non-verbale intelligentie. Nagegaan zal worden in hoeverre de verwachtingen, die in dit verband zijn geformuleerd en die betrekking hebben op zowel de convergente als de discriminante validiteit van de ontwikkelde tests, juist zijn. Bovendien zal worden nagegaan welke samenhang er bestaat tussen de scores op de ontwikkelde tests en rapportcijfers voor lezen, taal en rekenen. Vooruitlopend op de vraag naar de realiteitswaarde van de onderscheiding van de drie geselecteerde leestaken wordt met behulp van multiple regressie-analyse onderzocht in hoeverre er ruimte bestaat voor de veronderstelling dat de drie taakgerichte tests een specifiek aspect van begrijpend lezen meten.

De resultaten van dit validiteitsonderzoek geven een indicatie van de validiteit van de ontwikkelde tests. Vanwege het ontbreken van een coherente en enigszins getoetste theorie op het gebied van begrijpend lezen kan uiteraard geen sprake zijn van nomologische validatie (vgl. Drenth, 1972). Het laatste betekent dat een afdoend antwoord op de vraag wat de verschillende tests meten in het kader van dit onderzoek niet verwacht kan worden.

Homogeniteit van de tests

De eerste vraag die in verband met de validiteit van de tests gesteld wordt, heeft betrekking op de homogeniteit van de tests. Deze vraag, die ook in de vorige paragraaf aan de orde was, zal nu worden beantwoord aan de hand van de resultaten van factoranalyses.

In het vooronderzoek werd een factoranalyse uitgevoerd op de Pearson-correlaties tussen de items van elke taakgerichte test. Uit deze factoranalyses

bleek dat zowel bij de Hoofdgedachtetest als bij de Bedoelingtest sprake was van één factor. Factoranalyse op de items van de Redeneertest suggereerde twee factoren, die als Onzeker en Zeker konden worden geïnterpreteerd (vgl. par. 4.4.).

Op de items van de vijf ontwikkelde tests, die in de voormeting (vorm A) waren afgenomen, werd ook nu factoranalyse uitgevoerd^x. Op de tetrachorische correlaties tussen de items van elke taakgerichte test werd een factoranalyse uitgevoerd volgens de hoofdassenmethode (met enen op de diagonaal). Volgens dezelfde methode werd een factoranalyse uitgevoerd op de Pearson-correlaties tussen de items van de FRCT. Bij de MCCT werd een factoranalyse uitgevoerd op de Pearson-correlaties tussen de subtests. Bij deze analyses werd het aantal factoren op maximaal vier gesteld. Tabel 5.6. geeft de eigenwaarden van de eerste vier factoren van elke test (inclusief de subtests van de Redeneertest) met tussen haakjes het percentage verklaarde variantie.

Tabel 5.6. Resultaten van de factoranalyses op de vijf tests voor begrijpend lezen (vorm A; voormeting: alle scholen). N = ca. 500.

	Eigenwaarden en percentages verklaarde variantie			
	1e Factor	2e Factor	3e Factor	4e Factor
HGD	11.1 (38.2)	1.5 (5.2)	1.4 (4.7)	1.2 (4.3)
BED	7.7 (32.2)	1.4 (5.8)	1.3 (5.4)	1.1 (4.7)
RED	8.5 (21.2)	5.3 (13.1)	1.8 (4.6)	1.8 (4.5)
Onzeker	6.7 (41.9)	1.2 (7.6)	1.0 (6.1)	1.0 (6.1)
Zeker	6.6 (27.5)	1.8 (7.5)	1.5 (6.4)	1.3 (5.4)
FRCT	6.9 (13.8)	2.0 (3.9)	1.6 (3.3)	1.5 (3.0)
MCCT	4.1 (51.1)	0.9 (10.6)	0.6 (8.1)	0.6 (7.0)

^x

Opgemerkt moet worden dat de resultaten van factoranalyses op 0 - 1 scores met de nodige voorzichtigheid moeten worden gehanteerd (vgl. Lord & Novick, 1974).

Uit tabel 5.6. blijkt dat zowel bij de Hoofdgedachtetest (vorm A) als bij de Bedoelingtest (vorm A) sprake is van één factor en dat de restfactoren van weinig betekenis zijn. Op grond van deze gegevens mogen deze tests als voldoende homogeen worden beschouwd.

De eigenwaarde en het percentage verklaarde variantie van de eerste en tweede factor van de Redeneertest (vorm A) indiceren dat bij deze test sprake is van twee factoren. In het vooronderzoek kon factor 1 in een varimax geroteerde twee-factorenoplossing geïnterpreteerd worden als Onzeker en factor 2 als Zeker. Varimax rotatie van de beide eerste factoren leverde ook nu hetzelfde resultaat op als in het vooronderzoek. Uit de tabel blijkt dat zowel bij de subtest Onzeker (vorm A) als bij de subtest Zeker (vorm A) sprake is van één factor. Op grond van deze gegevens mogen deze subtests als voldoende homogeen worden beschouwd.

Uit tabel 5.6. blijkt verder dat het percentage variantie dat bij de FRCT (vorm A) door de eerste factor wordt verklaard aan de lage kant is. Bij deze test moest gekozen worden voor Pearson-correlaties in plaats van tetrachorische correlaties vanwege het optreden van Heywood-cases (variabelen i.c. items met communaliteiten hoger dan 1.0). Gezien het feit dat de FRCT de maximale range van de moeilijkheidsgraad van de items (van 0 tot 1) omvat, hetgeen een sterke invloed heeft op de intercorrelaties tussen de items vanwege de afhankelijkheid van de itemvarianties van itemmoeilijkheidsgraden, en het feit dat de items van deze test experimenteel niet onafhankelijk zijn, kan de vraag gesteld worden of het wel zinvol is om de homogeniteit van deze test met behulp van factoranalyse te bepalen.

Uit de tabel blijkt dat bij de MCCT (vorm A) sprake is van één factor; deze test kan ook als voldoende homogeen worden beschouwd.

De bovengenoemde factoranalyses werden ook uitgevoerd op de items c.q. subtests van de vijf ontwikkelde tests, die in de nameting (vorm B) en re-tentiemeting (vorm A) werden afgenomen op de controlescholen. Tabel 5.7. en 5.8. geven de resultaten van deze factoranalyses.

Uit tabel 5.7. blijkt dat de eigenwaarden van de verschillende factoren bij de B-vormen van de tests eenzelfde verloop vertonen als bij de A-vormen. De gegevens van tabel 5.8. bevestigen wat reeds eerder werd geconstateerd naar aanleiding van de factoranalyses op de tests bij de voormeting (vgl. tabel 5.6.). Zoals reeds eerder vermeld, worden de tests over de meetmomen-

ten heen over het algemeen - een uitzondering vormt de FRCT - homogener. Dit blijkt uit de stijging van het percentage variantie dat door de eerste factor wordt verklaard.

Tabel 5.7. Resultaten van de factoranalyses op de vijf tests voor begrijpend lezen (vorm B; nameting: controlescholen 5, 6, 7, 8).
N = ca. 265.

		Eigenwaarden en percentages verklaarde variantie			
		1e Factor	2e Factor	3e Factor	4e Factor
HGD		11.7 (40.4)	1.7 (5.9)	1.5 (5.1)	1.3 (4.5)
BED		8.9 (37.2)	1.7 (7.0)	1.5 (6.1)	1.3 (5.5)
RED		11.6 (29.0)	3.7 (9.2)	1.9 (4.7)	1.8 (4.6)
	Onzeker	6.9 (43.3)	1.2 (7.6)	1.1 (6.8)	1.0 (6.2)
	Zeker	7.7 (32.1)	1.7 (7.1)	1.5 (6.0)	1.4 (5.8)
FRCT		6.4 (12.9)	2.0 (3.9)	1.8 (3.6)	1.7 (3.3)
MCCT		4.6 (58.1)	0.7 (8.9)	0.6 (7.7)	0.5 (6.7)

Tabel 5.8. Resultaten van de factoranalyses op de vijf tests voor begrijpend lezen (vorm A; retentiemeting: controlescholen 6, 7, 8).
N = ca. 220.

		Eigenwaarden en percentages verklaarde variantie			
		1e Factor	2e Factor	3e Factor	4e Factor
HGD		12.1 (41.6)	1.7 (5.7)	1.6 (5.6)	1.4 (4.8)
BED		10.6 (44.0)	2.0 (8.3)	1.4 (6.1)	1.2 (5.0)
RED		13.1 (32.7)	3.4 (8.6)	2.5 (6.4)	2.0 (5.1)
	Onzeker	6.8 (42.7)	1.3 (8.3)	1.2 (7.4)	1.1 (6.7)
	Zeker	9.0 (37.3)	1.9 (7.9)	1.7 (7.0)	1.6 (6.9)
FRCT		6.6 (13.3)	2.2 (4.4)	2.0 (3.9)	1.8 (3.7)
MCCT		4.9 (61.5)	0.6 (8.0)	0.6 (7.8)	0.4 (5.4)

Het gevondene kan als volgt worden samengevat. De Hoofdgedachtetest, de Bedoelingtest en de MCCT meten één factor. De Redeneertest bevat twee interpreteerbare factoren. Over de homogeniteit van de FRCT kan met behulp van factoranalyse waarschijnlijk geen zinvolle uitspraak worden gedaan.

Samenhang tussen de vijf tests

De vraag die in verband met de validiteit van de tests zeker beantwoord moet worden, is de vraag naar de samenhang tussen de vijf tests. Wat betreft de samenhang tussen de drie taakgerichte tests verwachten we dat deze tests in aanzienlijke^{*} mate zullen samenhangen. Deze verwachting is gebaseerd op de veronderstelling dat elk van deze tests in sterke mate een beroep doet op hogere cognitieve processen i.c. inferentieprocessen. Op basis van dezelfde veronderstelling verwachten we dat de FRCT en de MCCT in aanzienlijke mate zullen samenhangen met de taakgerichte tests. Als we uitgaan van de veronderstelling van O'Reilly c.s. dat de MCCT voor een groot deel een maat is voor letterlijk begrijpen, dan moet de MCCT middelmatig tot laag correleren met de taakgerichte tests en de FRCT (vgl. par. 3.3.2.).

In tabel 5.9. worden de product-moment correlatiecoëfficiënten weergegeven tussen de drie taakgerichte tests (inclusief de subtests voor Red.), die bij de voormeting (vorm A) werden afgenomen.

Tabel 5.9. Correlaties tussen de drie taakgerichte tests voor begrijpend lezen (vorm A; voormeting: alle scholen). N = ca. 500.

	HGD	BED
BED	.70	-
RED	.60	.56
Onzeker	.60	.55
Zeker	.35	.35

^{*} Een correlatie van .40 tot .60 wordt met de term middelmatig aangeduid. Een correlatie van .60 tot .70 wordt aanzienlijk genoemd. Een correlatie beneden .40 wordt gekarakteriseerd als laag of vrij laag, terwijl een correlatie boven .70 als hoog of vrij hoog wordt aangeduid.

Uit tabel 5.9. blijkt dat er een aanzienlijke samenhang bestaat tussen de Hoofdgedachtetest (vorm A) en de Bedoelingtest (vorm A) en tussen de Hoofdgedachtetest en de Redeneertest (vorm A). De samenhang tussen de Bedoelingtest en de Redeneertest is middelmatig en niet in overeenstemming met de verwachting. De subtest Onzeker correleert even hoog met de twee andere taakgerichte tests als de Redeneertest in zijn geheel. De subtest Zeker heeft een lage samenhang met de andere taakgerichte tests.

In tabel 5.10. worden de correlaties weergegeven tussen de drie taakgerichte tests (inclusief de subtests voor Red.) en de cloze tests die eveneens bij de voormeting werden afgenomen. De correlatie tussen de FRCT (vorm A) en de MCCT (vorm A) bedraagt .68.

Tabel 5.10. Correlaties tussen de drie taakgerichte tests en de twee cloze tests (vorm A; voormeting: alle scholen). N = ca. 500.

	FRCT	MCCT
HGD	.59	.56
BED	.61	.54
RED	.48	.45
Onzeker	.48	.44
Zeker	.28	.27

Uit tabel 5.10. blijkt dat de correlaties, die tijdens de voormeting werden gevonden niet in overeenstemming zijn met de verwachting. De samenhang tussen de taakgerichte tests en de cloze tests is in het algemeen middelmatig.

Ter verificatie van de bovenstaande gegevens werden de onderlinge correlaties berekend tussen de vijf tests (inclusief de subtests voor Red.), die tijdens de voormeting (vorm A), de nameting (vorm B) en de retentiemeting (vorm A) werden afgenomen op drie controlescholen. De tabellen 5.11, 5.12. en 5.13. geven een overzicht van de gevonden correlaties tussen de totaalscores op de vijf tests voor begrijpend lezen.

Tabel 5.11. Correlaties tussen de vijf tests voor begrijpend lezen.
(vorm A; voormeting: controlescholen 6, 7, 8). N = ca. 220.

	1	2	3	4	5	6	7
1. HGD	-						
2. BED	.67	-					
3. RED	.59	.48	-				
4. Onzeker	.63	.51	-	-			
5. Zeker	.35	.26	-	.21	-		
6. FRCT	.57	.55	.47	.43	.31	-	
7. MCCT	.55	.45	.45	.42	.29	.63	-

Tabel 5.12. Correlaties tussen de vijf tests voor begrijpend lezen.
(vorm B; nameting: controlescholen 6, 7, 8). N = ca. 220.

	1	2	3	4	5	6	7
1. HGD	-						
2. BED	.75	-					
3. RED	.72	.68	-				
4. Onzeker	.63	.59	-	-			
5. Zeker	.62	.58	-	.50	-		
6. FRCT	.62	.65	.65	.56	.58	-	
7. MCCT	.71	.73	.66	.55	.59	.71	-

Tabel 5.13. Correlaties tussen de vijf tests voor begrijpend lezen.
(vorm A; retentiemeting: controlescholen 6, 7, 8). N = ca. 220.

	1	2	3	4	5	6	7
1. HGD	-						
2. BED	.72	-					
3. RED	.75	.66	-				
4. Onzeker	.67	.59	-	-			
5. Zeker	.66	.58	-	.55	-		
6. FRCT	.61	.55	.56	.49	.50	-	
7. MCCT	.68	.69	.67	.56	.61	.60	-

Uit de tabellen 5.11., 5.12. en 5.13. blijkt dat de samenhang tussen de drie taakgerichte tests bij de nameting en retentiemeting beduidend hoger is dan bij de voormeting en in overeenstemming met de eerder geformuleerde verwachting. Deze hogere samenhang geldt met name voor de Redeneertest met de twee andere taakgerichte tests. Bij de nameting en retentiemeting bestaat er een aanzienlijke tot vrij hoge samenhang tussen de Redeneertest en de twee andere taakgerichte tests. Een ander gegeven is dat de subtest Zeker bij de nameting en retentiemeting even hoog correleert met de andere tests als de subtest Onzeker. Uit deze tabellen blijkt verder dat de samenhang tussen de drie taakgerichte tests en de FRCT bij de nameting en retentiemeting varieert van middelmatig tot aanzienlijk. De samenhang tussen de drie taakgerichte tests en de MCCT is in de nameting en retentiemeting aanzienlijk tot hoog. Deze correlaties komen voor een groot deel overeen met de eerder geformuleerde verwachting en zijn wat betreft de MCCT in strijd met de veronderstelling van O'Reilly c.s. (vgl. par. 3.3.2.).

Het feit dat in de nameting en retentiemeting de correlaties tussen de drie taakgerichte tests onderling en tussen de taakgerichte tests en de cloze tests hoger zijn dan in de voormeting, is moeilijk te verklaren. Een voor de hand liggende verklaring zou zijn dat de variantie tussen de vooren nameting respectievelijk retentiemeting is toegenomen. De verschillen tussen de correlaties van de tests bij de voormeting en nameting zijn echter te groot om deze volledig toe te schrijven aan de vrij geringe toename van de variantie (vgl. par. 5.1.3. en 5.1.5.). Een mogelijke aanvullende verklaring kan zijn dat de vijf ontwikkelde tests tijdens de voormeting onder minder gunstige omstandigheden zijn afgenomen dan tijdens de nameting en retentiemeting. Bij de voormeting werden in totaal tien tests afgenomen waaronder de vijf ontwikkelde tests als laatste. De mogelijkheid is niet uitgesloten dat bij een aantal leerlingen de motivatie parten heeft gespeeld bij het maken van de laatste tests. Bovendien kan het begrijpen van de taak tijdens de voormeting voor sommige leerlingen moeilijkheden hebben opgeleverd.

Een ander gegeven is de lage samenhang van de subtest Zeker met de andere tests bij de voormeting. Zoals uit par. 5.1.3. blijkt, kan deze lage correlatie niet worden toegeschreven aan de betrouwbaarheid of de spreiding van deze subtest. De moeilijkheidsgraad van deze subtest vormt evenmin een verklaring; de subtest Onzeker is moeilijker dan de subtest Zeker.

Tegen de eerder geformuleerde verwachting betreffende de samenhang tussen de taakgerichte tests en de MCCT kan de vraag gesteld worden of op grond van de correlaties tussen de totaalscores van de tests een structuur te construeren is met twee dimensies die benoemd zouden kunnen worden als inferentieel begrijpend lezen, gemeten met de taakgerichte tests en de FRCT én letterlijk begrijpend lezen, gemeten met de MCCT. Om deze vraag te beantwoorden werd een factoranalyse uitgevoerd volgens de hoofdasenmethode (met enen op de diagonaal). Er werden twee factoren geëxtraheerd die met behulp van het varimax-criterium zijn geroteerd. De resultaten van de factoranalyses zijn opgenomen in tabel 5.14. en tabel 5.15.

In de analyse die betrekking heeft op de voormeting (vorm A; alle scholen) zijn de factoren in één richting te interpreteren nl. één factor waarop de taakgerichte tests hoog laden en één factor waarop de cloze tests hoog laden. Hetzelfde geldt voor de analyse die betrekking heeft op de voormeting van de drie controlescholen. In de analyses die betrekking hebben op de nameting en retentiemeting zien we dat de MCCT hoog gaat laden op de eerste factor. Op grond van deze en reeds eerder vermelde gegevens kan niet worden geconcludeerd dat de taakgerichte tests en de MCCT twee verschillende aspecten van begrijpend lezen meten.

Het voorgaande kan als volgt worden samengevat. Er bestaat een aanzienlijke tot vrij hoge samenhang tussen de Hoofdgedachtetest en de Bedoelingstest. De samenhang tussen deze tests en de Redeneertest varieert in de voormeting van middelmatig tot aanzienlijk en in de nameting en retentiemeting op drie controlescholen van aanzienlijk tot vrij hoog. De samenhang tussen de drie taakgerichte tests en de cloze tests is bij de voormeting - tegen de verwachting - middelmatig. Bij de nameting en retentiemeting is de samenhang tussen de taakgerichte tests en de MCCT aanzienlijk tot vrij hoog, terwijl de samenhang tussen de taakgerichte tests en de FRCT varieert van middelmatig tot aanzienlijk. Op grond van onder andere de resultaten van de factoranalyses kan niet worden geconcludeerd dat de taakgerichte tests en de MCCT twee verschillende aspecten van begrijpend lezen meten.

Tabel 5.14. Ladingen op varimax-geroteerde principale componenten van de vijf tests voor begrijpend lezen. (vorm A; voormeting: alle scholen). N = ca. 500.

	Factor 1	Factor 2
HGD	.74	.46
BED	.71	.47
RED	.88	.16
FRCT	.34	.83
MCCT	.25	.88

Tabel 5.15. Ladingen op varimax-geroteerde principale componenten van de vijf tests voor begrijpend lezen. (vorm A - voormeting; vorm B - nameting; vorm A - retentiemeting: controlescholen 6, 7, 8). N = ca. 220.

		Factor 1	Factor 2
voor- meting	HGD	.79	.40
	BED	.76	.33
	RED	.82	.20
	FRCT	.37	.81
	MCCT	.25	.88
na- meting	HGD	.88	.31
	BED	.79	.42
	RED	.76	.42
	FRCT	.34	.91
	MCCT	.59	.67
retentie- meting	HGD	.83	.35
	BED	.84	.26
	RED	.84	.28
	FRCT	.34	.94
	MCCT	.75	.42

Samenhang met technisch lezen

Een belangrijke vraag in verband met de validiteit van de tests is de vraag naar de samenhang van deze tests met technisch lezen. Als de tests inderdaad begrijpend lezen meten, mag verwacht worden dat in het vierde leerjaar een vrij lage samenhang wordt gevonden tussen deze tests en technisch lezen. Verwacht wordt dat de drie taakgerichte tests waarschijnlijk lager correleren met technisch lezen dan de cloze tests.

Zoals reeds eerder vermeld, werd in het onderzoek de Eén-Minuut-Test, de Schriftelijke Opdrachten en de Stilleestest afgenomen. De eerste test meet het technisch lezen, terwijl de twee andere tests bedoeld zijn het begrijpend lezen te meten. In het ijkingsonderzoek (vgl. Brus & Voeten, 1973) werd de samenhang tussen deze tests bepaald. Voor het vierde leerjaar bedroeg de correlatie tussen de Eén-Minuut-Test en de Schriftelijke Opdrachten .42 en tussen de Eén-Minuut-Test en de Stilleestest .42. Deze correlaties hebben betrekking op de afnames in februari-maart. Uit dit onderzoek bleek dat in het vierde leerjaar een vrij laag percentage ($\pm 23\%$) van de variantie in de scores op de tests voor begrijpend lezen verklaard wordt door de lineaire relatie met de Eén-Minuut-Test en omgekeerd.

In ons onderzoek bedraagt de correlatie tussen de Eén-Minuut-Test en de Schriftelijke Opdrachten voor het vierde leerjaar .51. De correlatie tussen de Eén-Minuut-Test en de Stilleestest bedraagt .38. Uit deze gegevens blijkt dat de correlatie tussen de Eén-Minuut-Test en de Stilleestest bijna gelijk is aan de correlatie die in het ijkingsonderzoek werd gevonden, terwijl de correlatie tussen de Eén-Minuut-Test en de Schriftelijke Opdrachten in het ijkingsonderzoek lager is.

In tabel 5.16. worden de correlaties weergegeven tussen de Eén-Minuut-Test en de vijf tests (inclusief de subtests voor Red.), die tijdens de voormeting werden afgenomen. Uit deze tabel blijkt dat de drie taakgerichte tests laag correleren met de Eén-Minuut-Test. Opvallend is de lage correlatie tussen de Hoofdgedachtetest (vorm A) respectievelijk de Redeneertest (vorm A) en de Eén-Minuut-Test. De subtest Zeker (vorm A) correleert zeer laag met de Eén-Minuut-Test. De cloze tests vertonen een middelmatige samenhang met technisch lezen; ze correleren hoger met technisch lezen dan de drie taakgerichte tests. Bovenstaande gegevens geven steun aan de verwachting dat de vijf tests voor begrijpend lezen voor een niet onbelangrijk

deel iets anders meten dan technisch lezen.

Tabel 5.16. Correlaties tussen de Eén-Minuut-Test en de vijf tests voor begrijpend lezen. (vorm A; voormeting: alle scholen).
N = ca. 500.

Eén-Minuut-Test	
HGD	.32
BED	.40
RED	.25
Onzeker	.24
Zeker	.15
FRCT	.48
MCCT	.50

Ter verificatie van het bovenstaande werden ook nu de correlaties berekend tussen de Eén-Minuut-Test en de vijf tests (inclusief de subtests voor Red.), die tijdens de voormeting (vorm A), de nameting (vorm B) en de retentiemeting (vorm A) werden afgenomen op drie controlescholen. Tabel 5.17. geeft een overzicht van de correlaties tussen de Eén-Minuut-Test - afgenomen tijdens de voormeting - en de vijf tests voor begrijpend lezen op drie verschillende meetmomenten.

Uit deze tabel blijkt dat de samenhang tussen de drie taakgerichte tests en de Eén-Minuut-Test bij de nameting en retentiemeting hoger is dan bij de voormeting; de samenhang blijft bij de drie meetmomenten echter laag. De subtest Zeker correleert bij de nameting en retentiemeting hoger met de Eén-Minuut-Test dan bij de voormeting. Zoals reeds eerder vermeld is het moeilijk de hogere correlaties in de na- en retentiemeting te verklaren.

De cloze tests vertonen ook bij de nameting en retentiemeting een middelmatige samenhang met technisch lezen. Ze correleren hoger met technisch lezen dan de drie taakgerichte tests, al is het verschil niet meer zo groot als bij de voormeting.

Tabel 5.17. Correlaties van de gevonden scores tussen de Eén-Minuut-Test en de vijf tests voor begrijpend lezen. (vorm A - voormeting; vorm B - nameting; vorm A - retentiemeting: controlescholen 6, 7, 8). N = ca. 220.

voormeting		nameting		retentiemeting	
Vorm A	Eén-Minuut-Test	Vorm B	Eén-Minuut-Test	Vorm A	Eén-Minuut-Test
HGD	.28	HGD	.39	HGD	.38
BED	.33	BED	.36	BED	.39
RED	.25	RED	.32	RED	.32
Onzeker	.22	Onzeker	.23	Onzeker	.25
Zeker	.18	Zeker	.32	Zeker	.32
FRCT	.46	FRCT	.42	FRCT	.42
MCCT	.45	MCCT	.44	MCCT	.46

Het bovenstaande kunnen we als volgt samenvatten. De drie taakgerichte tests correleren - overeenkomstig de verwachting - laag met de Eén-Minuut-Test. De cloze tests vertonen een middelmatige samenhang met technisch lezen.

Samenhang met begrijpend lezen

Een vraag die in het kader van het validiteitsonderzoek zeker beantwoord moet worden, is de vraag naar de samenhang van de ontwikkelde tests met andere tests voor begrijpend lezen. Verwacht wordt dat de vijf tests voor begrijpend lezen een aanzienlijke samenhang vertonen met tests voor begrijpend lezen in het algemeen.

Zoals reeds eerder vermeld, werden in het onderzoek twee gestandaardiseerde tests voor begrijpend lezen afgenomen nl. de Schriftelijke Opdrachten en de Stilleestest, beide voor het vierde leerjaar. De eerste test heeft betrekking op het begrijpen van de betekenis van korte aanwijzingen gevat in één of twee zinnen. De tweede test heeft betrekking op het begrijpen van de betekenis van langere teksten. Opgemerkt moet worden dat bij de Stilleestest de items die betrekking hebben op een zelfde tekst experimen-

teel niet onafhankelijk zijn. Bovendien bestaan er twijfels over de betrouwbaarheid van deze test (vgl. Danhof, Smelt & Van der Wissel, 1978).

In het kader van het ijkingsonderzoek (vgl. Brus & Van Bergen, 1973) werd de samenhang bepaald tussen de Schriftelijke Opdrachten en de Stilleestest. Voor het vierde leerjaar bedroeg de correlatie tussen beide tests .67. Uit dit gegeven blijkt dat er een aanzienlijke samenhang bestaat tussen deze twee tests voor begrijpend lezen. In ons onderzoek bedraagt de correlatie tussen de Schriftelijke Opdrachten en de Stilleestest voor het vierde leerjaar .62. De KR-20 betrouwbaarheidscoëfficiënt van de Schriftelijke Opdrachten bedraagt .87; dezelfde coëfficiënt bereikt voor de Stilleestest een waarde van .78.

In tabel 5.18. worden de correlaties weergegeven tussen de Stilleestest respectievelijk Schriftelijke Opdrachten en de vijf tests (inclusief de subtests voor Red.), die tijdens de voormeting werden afgenomen.

Tabel 5.18. Correlaties tussen Stilleestest/Schriftelijke Opdrachten en de vijf tests voor begrijpend lezen. (vorm A; voormeting: alle scholen). N = ca. 500.

	Stilleestest	Schr. Opdr.
HGD	.59	.61
BED	.66	.62
RED	.50	.53
Onzeker	.51	.50
Zeker	.29	.33
FRCT	.60	.74
MCCT	.52	.68

Uit deze tabel blijkt dat de drie taakgerichte tests middelmatig tot aanzienlijk correleren met de Stilleestest en de Schriftelijke Opdrachten. De subtest Zeker (vorm A) correleert laag met de twee geijkte tests voor begrijpend lezen. Verder blijkt dat de cloze tests in het algemeen een aanzienlijke samenhang vertonen met de Stilleestest en de Schriftelijke Opdrachten.

Ter verificatie van het bovenstaande werden ook nu de correlaties berekend tussen de Stilleestest respectievelijk Schriftelijke Opdrachten en de vijf tests (inclusief de subtests voor Red.), die tijdens de voormeting (vorm A), de nameting (vorm B) en de retentiemeting (vorm A) werden afgenomen op drie controlescholen. De tabellen 5.19. en 5.20 geven een overzicht van de correlaties tussen de Stilleestest respectievelijk Schriftelijke Opdrachten en de vijf tests voor begrijpend lezen.

Tabel 5.19. Correlaties van de gevonden scores tussen de Stilleestest en de vijf tests voor begrijpend lezen (vorm A - voormeting; vorm B - nameting; vorm A - retentiemeting; controlescholen 6, 7, 8). N = ca. 220.

voormeting		nameting		retentiemeting	
Vorm A	Stilleestest	Vorm B	Stilleestest	Vorm A	Stilleestest
HGD	.52	HGD	.51	HGD	.53
BED	.63	BED	.53	BED	.49
RED	.46	RED	.51	RED	.46
Onzeker	.48	Onzeker	.53	Onzeker	.49
Zeker	.26	Zeker	.37	Zeker	.34
FRCT	.59	FRCT	.59	FRCT	.50
MCCT	.44	MCCT	.52	MCCT	.47

Tabel 5.20. Correlaties van de gevonden scores tussen de Schriftelijke Opdrachten en de vijf tests voor begrijpend lezen. (vorm A - voormeting; vorm B - nameting; vorm A - retentiemeting; controlescholen 6, 7, 8). N = ca. 220.

voormeting		nameting		retentiemeting	
Vorm A	Schr. Opdr.	Vorm B	Schr. Opdr.	Vorm A	Schr. Opdr.
HGD	.55	HGD	.56	HGD	.61
BED	.52	BED	.62	BED	.64
RED	.54	RED	.62	RED	.61
Onzeker	.48	Onzeker	.56	Onzeker	.54
Zeker	.38	Zeker	.51	Zeker	.54
FRCT	.70	FRCT	.72	FRCT	.68
MCCT	.66	MCCT	.64	MCCT	.64

Uit tabel 5.19. blijkt dat de samenhang tussen de drie taakgerichte tests en de Stilleestest bij de drie meetmomenten in het algemeen middelmatig van aard is. Opvallend is dat de subtest Zeker ook in de nameting en retentiemeting laag correleert met de Stilleestest. De cloze tests vertonen eveneens een middelmatige samenhang met de Stilleestest. Uit tabel 5.20. blijkt dat de drie taakgerichte tests bij de nameting en retentiemeting over het algemeen aanzienlijk correleren met de Schriftelijke Opdrachten. De cloze tests vertonen op de drie meetmomenten een aanzienlijke tot hoge samenhang met de Schriftelijke Opdrachten.

Uit het bovenstaande blijkt dat de vijf ontwikkelde tests tegen de verwachting in slechts een middelmatige samenhang vertonen met de Stilleestest. Een verklaring voor deze middelmatige samenhang zou gevonden kunnen worden in het feit dat de Stilleestest niet voldoende betrouwbaar is. Overeenkomstig de verwachting correleren de vijf tests over het algemeen in aanzienlijke mate met de Schriftelijke Opdrachten.

Samenhang met woordenschat

Een volgende vraag die in verband met de validiteit van de tests gesteld wordt, is de vraag naar de samenhang van de ontwikkelde tests met woordenschat. Verwacht wordt dat de tests een aanzienlijke samenhang vertonen met een test die woordenschat meet. Uit de theorievorming en het onderzoek op het gebied van lezen is bekend dat woordenschat een belangrijk aspect vormt van begrijpend lezen.

In het onderzoek werd de Woordenschattest van Stijnen (1975) afgenomen. Deze test meet de leeswoordenschat, die ook als een aspect van de taalvaardigheid in het algemeen beschouwd kan worden. In het kader van het ijkingsonderzoek (vgl. Brus & Van Bergen, 1973; Stijnen, 1975) werd de samenhang bepaald tussen de Woordenschattest en twee tests voor begrijpend lezen nl. de Schriftelijke Opdrachten en de Stilleestest. Voor het vierde leerjaar bedroeg de correlatie tussen de Woordenschattest en Schriftelijke Opdrachten .76. De correlatie tussen de Woordenschattest en de Stilleestest bedroeg .65. Uit deze gegevens blijkt dat er een aanzienlijke tot vrij hoge samenhang bestaat tussen de Woordenschattest en de twee bovengenoemde tests voor begrijpend lezen.

In ons onderzoek bedraagt de correlatie tussen de Woordenschattest en

de Schriftelijke Opdrachten .77; de correlatie tussen de Woordenschattest en de Stilleestest is .67. De KR-20 betrouwbaarheidscoëfficiënt van de Woordenschattest bedroeg tijdens de voormeting .89.

Tabel 5.21. geeft een overzicht van de correlaties tussen de Woordenschattest en de vijf tests voor begrijpend lezen (inclusief de subtests voor Red.), die tijdens de voormeting werden afgenomen.

Tabel 5.21. Correlaties tussen de Woordenschattest en de vijf tests voor begrijpend lezen (vorm A; voormeting: alle scholen).
N = ca. 500.

	Woordenschattest
HGD	.71
BED	.69
RED	.57
Onzeker	.56
Zeker	.34
FRCT	.76
MCCT	.70

Uit tabel 5.21. blijkt dat er een aanzienlijke samenhang bestaat tussen de Woordenschattest en twee taakgerichte tests nl. de Hoofdgedachtetest en de Bedoelingtest. De cloze tests correleren vrij hoog met de Woordenschattest. Dit laatste geldt met name voor de FRCT.

Ter verificatie van het bovenstaande werden ook nu de correlaties berekend tussen de Woordenschattest en de vijf tests (inclusief de subtests voor Red.), die tijdens de voormeting (vorm A), de nameting (vorm B) en de retentiemeting (vorm A) werden afgenomen op drie controlescholen. Tabel 5.22. geeft een overzicht van de correlaties tussen de Woordenschattest en de vijf tests voor begrijpend lezen.

Tabel 5.22. Correlaties van de gevonden scores tussen de Woordenschattest en de vijf tests voor begrijpend lezen. (vorm A - voormeting; vorm B - nameting; vorm A - retentiemeting: controlescholen 6, 7, 8). N = ca. 220.

voormeting		nameting		retentiemeting	
Vorm A	Woorden-schattest	Vorm B	Woorden-schattest	Vorm A	Woorden-schattest
HGD	.67	HGD	.66	HGD	.69
BED	.63	BED	.66	BED	.63
RED	.55	RED	.65	RED	.65
Onzeker	.55	Onzeker	.61	Onzeker	.61
Zeker	.34	Zeker	.53	Zeker	.54
FRCT	.74	FRCT	.71	FRCT	.69
MCCT	.67	MCCT	.67	MCCT	.62

Uit tabel 5.22. blijkt dat er - met uitzondering van de Redeneertest in de voormeting - een aanzienlijke samenhang bestaat tussen de drie taakgerichte tests en de Woordenschattest. Uit de tabel blijkt verder dat de cloze tests aanzienlijk tot vrij hoog correleren met de Woordenschattest.

Het gevondene kan als volgt worden samengevat. De drie taakgerichte tests vertonen in het algemeen een aanzienlijke samenhang met de Woordenschattest. De samenhang tussen de cloze tests en de Woordenschattest varieert van aanzienlijk tot vrij hoog. Deze resultaten komen grotendeels overeen met de eerder geformuleerde verwachting.

Samenhang met non-verbale intelligentie

Een andere vraag die in verband met het validiteitsonderzoek van de tests gesteld wordt, heeft betrekking op de samenhang van deze tests met non-verbale intelligentie. Op basis van de bestudeerde onderzoeksliteratuur verwachten we een middelmatige tot vrij lage samenhang tussen de vijf ontwikkelde tests en een maat voor non-verbale intelligentie.

In ons onderzoek werd de Standard Progressive Matrices van Raven (1958) afgenomen. Deze test is bedoeld om de non-verbale intelligentie te meten.

"It is often useful to describe the scale as a test of observation and clear thinking. By itself it is not a test of "general intelligence" and it is always a mistake to describe it as such." (Raven, 1958, pag. 2). In tabel 5.23. worden de gevonden correlatiecoëfficiënten van de Standard Progressive Matrices én de Stilleestest, de Schriftelijke Opdrachten en de Woordenschattest weergegeven voor de betreffende steekproef. De KR-20 betrouwbaarheidscoëfficiënt van de Standard Progressive Matrices bedroeg tijdens de voormeting .91.

Tabel 5.23. Correlaties tussen de Standard Progressive Matrices en de Stilleestest, Schriftelijke Opdrachten en Woordenschattest. (voormeting: alle scholen). N = ca. 500.

	Standard Progressive Matrices
Stilleestest	.39
Schriftelijke Opdrachten	.49
Woordenschattest	.46

Uit tabel 5.23. blijkt dat er slechts een middelmatige samenhang bestaat tussen de Standard Progressive Matrices en de tests voor begrijpend lezen en woordenschat.

Tabel 5.24. geeft een overzicht van de correlaties van de Standard Progressive Matrices en de vijf tests voor begrijpend lezen (inclusief de subtests voor Red.), die tijdens de voormeting op alle scholen werden afgenomen.

Uit tabel 5.24. blijkt dat de drie taakgerichte tests middelmatig correleren met de Standard Progressive Matrices. De Redeneertest (vorm A) vertoont geen hogere samenhang met de betreffende intelligentietest dan de beide andere taakgerichte tests. De subtest Zeker correleert laag met de Standard Progressive Matrices. De cloze tests vertonen eveneens een middelmatige samenhang met non-verbale intelligentie.

Tabel 5.24. Correlaties tussen de Standard Progressive Matrices en de vijf tests voor begrijpend lezen. (vorm A; voormeting: alle scholen). N = ca. 500.

	Standard Progressive Matrices
HGD	.46
BED	.44
RED	.45
Onzeker	.43
Zeker	.28
FRCT	.41
MCCT	.40

Ter verificatie van het bovenstaande werden ook nu de correlaties berekend tussen de Standard Progressive Matrices en de vijf tests (inclusief de subtests voor Red.), die tijdens de voormeting (vorm A), de nameting (vorm B) en de retentiemeting (vorm A) werden afgenomen op drie controlescholen. Tabel 5.25. geeft een overzicht van de gevonden correlaties tussen de Standard Progressive Matrices - afgenomen tijdens de voormeting - en vijf tests voor begrijpend lezen op drie verschillende meetmomenten.

Tabel 5.25. Correlaties van de gevonden scores tussen de Standard Progressive Matrices en de vijf tests voor begrijpend lezen. (vorm A - voormeting; vorm B - nameting; vorm A - retentiemeting: controlescholen 6, 7, 8). N = ca. 220.

voormeting		nameting		retentiemeting	
Vorm A	Standard Progressive Matrices	Vorm B	Standard Progressive Matrices	Vorm A	Standard Progressive Matrices
HGD	.47	HGD	.46	HGD	.44
BED	.40	BED	.45	BED	.41
RED	.42	RED	.51	RED	.49
Onzeker	.46	Onzeker	.46	Onzeker	.48
Zeker	.23	Zeker	.43	Zeker	.39
FRCT	.42	FRCT	.44	FRCT	.34
MCCT	.35	MCCT	.47	MCCT	.43

Uit tabel 5.25. blijkt dat de drie taakgerichte tests een middelmatige samenhang vertonen met de Standard Progressive Matrices. De Redeneertest correleert in de nameting en retentiemeting iets hoger met de Standard Progressive Matrices dan de twee andere taakgerichte tests. De subtest Zeker blijkt in de nameting en retentiemeting hoger met de Standard Progressive Matrices te correleren dan in de voormeting. De cloze tests vertonen in het algemeen een middelmatige samenhang met de Standard Progressive Matrices.

Het gevondene kan als volgt worden samengevat. De drie taakgerichte tests vertonen evenals de cloze tests een middelmatige samenhang met een test voor non-verbale intelligentie. Deze resultaten komen grotendeels overeen met de eerder geformuleerde verwachting.

Samenhang met het praktijkoordeel

Een volgende vraag die in verband met de validiteit van de tests gesteld wordt, is de vraag naar de samenhang tussen de uitslag van de tests en het oordeel van de leerkrachten over de schoolprestaties van hun leerlingen. Tijdens het onderzoek werden de rapportcijfers (waarderingen) voor lezen, taal en rekenen genoteerd. Zowel de cijfers van het Kerst- als van het Paasrapport werden verzameld. Verwacht wordt dat de scores op de tests hoger correleren met het cijfer voor lezen dan met het cijfer voor taal en rekenen. Bovendien wordt verwacht dat de scores op de tests hoger correleren met het cijfer voor begrijpend lezen dan met het cijfer voor technisch lezen.

Bij de bepaling van de samenhang tussen testcores en praktijkoordelen doen zich enkele problemen voor. Een eerste probleem is dat rapporten van verschillende scholen vaak verschillende indelingen of onderverdelingen voor de diverse vakken of leergebieden bevatten. Ook in de rapporten van de scholen die bij het onderzoek betrokken waren, worden voor lezen, taal en rekenen niet steeds dezelfde aspecten of componenten onderscheiden. Zo wordt in sommige rapporten geen onderscheid gemaakt tussen technisch en begrijpend lezen. Een tweede probleem is dat scholen vaak verschillende manieren hanteren om de waarderingen in de rapporten weer te geven. Sommige scholen geven de voorkeur aan waarderingen in cijfers, terwijl andere scholen verbale kwalificaties geven. In ons geval werd in het merendeel van de scholen met cijferrapporten gewerkt. Een derde probleem is dat de cijfers

of waarderingen van verschillende leerkrachten niet onderling vergelijkbaar zijn. Elke leerkracht kiest een eigen niveau bij het geven van cijfers en hanteert waarschijnlijk andere criteria dan een collega.

Wat betreft het eerste probleem - het verschillend aantal aspecten of componenten voor lezen, taal en rekenen - werd geen poging gedaan om verschillende aspecten onder één noemer te brengen en op die manier één cijfer of waardering voor lezen, taal en rekenen samen te stellen. De voorkeur werd gegeven aan die aspecten voor lezen, taal en rekenen die in het merendeel van de rapporten voorkwamen. De meest voorkomende aspecten waren: technisch en begrijpend lezen; schriftelijk taalgebruik, mondeling taalgebruik en dictee; hoofdrekenen, cijferen en redactierekenen. Deze keuze betekende dat bij de berekening van de correlaties tussen de testcores en de praktijkoordelen niet alle klassen (19) bij elk aspect van lezen, taal en rekenen konden worden betrokken. In de tabellen zal steeds het aantal klassen worden weergegeven waarvoor de correlaties zijn berekend.

Wat betreft het tweede probleem - de verschillende manieren waarop waarderingen in rapporten worden weergegeven - werd als volgt te werk gegaan. Na overleg met de betrokken leerkrachten werden de verbale kwalificaties op een zodanige wijze in cijfers omgezet dat de rangordening in de kwalificaties behouden bleef. Een verbale omschrijving als voldoende/ruim voldoende werd als 650 gecodeerd; nuances in de cijfers zoals 6+, 6½, 7- werden gecodeerd als 625, 650 en 675.

Wat betreft het derde probleem - de onvergelykbaarheid van de cijfers of waarderingen van verschillende leerkrachten onderling - werd als volgt gehandeld. Voor iedere leerkracht (d.w.z. klas) afzonderlijk werden de correlaties bepaald tussen de rapportcijfers en de testcores. Als maat voor centrale tendentie voor de op deze wijze berekende correlaties werd de mediane waarde genomen. Een correctie voor de "restriction of range" werd niet berekend.

Hoewel de rapportcijfers als ordinale gegevens dienen te worden opgevat, zijn bij het bepalen van de samenhang tussen rapportcijfers en testcores Pearson-correlaties berekend. Uit een vergelijking van de mediane correlaties tussen de rapportcijfers onderling bleek namelijk dat de numerieke waarden van de Spearman- en de Pearson-correlaties niet veel verschilden.

In tabel 5.26. worden de mediane correlaties (volgens Spearman) tussen de rapportcijfers met Kerstmis weergegeven.

Tabel 5.26. Correlaties tussen de rapportcijfers met Kerstmis: medianen van de correlaties per klas (het aantal klassen tussen haakjes).

	Techn. lezen (14)	Begrijp. lezen (16)	Schrift. taalgebr. (16)	Mond. taalgebr. (16)	Dictee (17)	Hoofd- rekenen (15)	Cijferen (15)	Redactie rekenen (10)
Techn. lezen								
Begrijp. Lezen	.56							
Schrift. taalgebr.	.58	.58						
Mond. taalgebr.	.59	.57	.58					
Dictee	.55	.51	.70	.48				
Hoofd- rekenen	.29	.42	.40	.34	.44			
Cijferen	.38	.49	.46	.46	.49	.78		
Redactie rekenen	.30	.45	.44	.35	.35	.77	.74	

Uit deze tabel blijkt dat de rapportcijfers voor lezen en taal over het algemeen een middelmatige samenhang vertonen. De cijfers voor lezen vertonen een hogere samenhang met de taalcijfers dan met de rekencijfers. De rapportcijfers voor hoofdrekenen, cijferen en redactierekenen correleren vrij hoog met elkaar. De mediane correlaties tussen de rapportcijfers van Pasen zijn over het algemeen hoger dan die van Kerstmis; ze vertonen verder hetzelfde beeld als weergegeven in tabel 5.26.

In tabel 5.27. worden de mediane correlaties weergegeven tussen de scores van de leerlingen op verschillende tests die in de voormeting werden

Tabel 5.27. Correlaties tussen enkele tests die betrekking hebben op technisch lezen, begrijpend lezen, woordenschat en non-verbale intelligentie én praktijkoordelen voor lezen, taal en rekenen: medianen van de correlaties per klas (het aantal klassen tussen haakjes).

	Leescijfer Kerstrapport		Taalcijfer Kerstrapport			Rekencijfer Kerstrapport		
	Technisch lezen (14)	Begrijpend lezen (16)	Schrift. taalgebr. (16)	Mond. taalgebr. (16)	Dictee (17)	Hoofd- rekenen (15)	Cijferen (15)	Redactie rekenen (10)
Eén-Minuut-Test	.69	.50	.52	.34	.62	.26	.23	.26
Stilleestest	.45	.53	.40	.37	.44	.36	.37	.34
Schriftelijke Opdrachten	.50	.64	.54	.50	.50	.57	.55	.50
Woordenschattest	.48	.60	.43	.48	.40	.45	.46	.47
Standard Progres- sive Matrices	.18	.44	.26	.26	.31	.48	.49	.48

afgenomen en de cijfers voor lezen, taal en rekenen van het Kerstrapport. Uit deze tabel blijkt dat er een aanzienlijke samenhang bestaat tussen de score op de Eén-Minuut-Test en het cijfer voor technisch lezen. In het ijkingsonderzoek (vgl. Brus & Voeten, 1973) werd een samenhang van .67 geconstateerd. Uit deze tabel blijkt verder dat de mediane correlaties tussen de scores op de tests voor begrijpend lezen en woordenschat (de Stillees-test, de Schriftelijke Opdrachten en de Woordenschattest) en het cijfer voor begrijpend lezen niet hoog zijn, maar in ieder geval wel hoger dan de mediane correlaties tussen deze scores en de cijfers voor taal en rekenen. In het ijkingsonderzoek (vgl. Brus & Van Bergen, 1973; Stijnen, 1975) werd een correlatie van .62 gevonden tussen de score op de Schriftelijke Opdrachten en het leescijfer en van .61 (gecorrigeerd voor "restriction of range") tussen de score op de Woordenschattest en het leescijfer.

In tabel 5.28. worden de mediane correlaties weergegeven tussen de scores van de leerlingen op de vijf ontwikkelde tests voor begrijpend lezen (inclusief de subtests voor Red.), die in de voormeting werden afgenomen en de cijfers voor lezen, taal en rekenen van het Kerstrapport. Uit deze tabel blijkt dat de samenhang tussen de drie taakgerichte tests (vorm A) en het cijfer voor begrijpend lezen middelmatig is, maar in ieder geval beduidend hoger is dan de samenhang met het cijfer voor technisch lezen. De mediane correlaties tussen de drie taakgerichte tests en het cijfer voor begrijpend lezen liggen ook hoger dan de correlaties met de taalcijfers. Hetzelfde geldt - met uitzondering van de Redeneertest - voor de correlaties met de cijfers voor rekenen. Opvallend is dat de Redeneertest hoger correleert met de cijfers voor rekenen dan voor taal. De subtest Zeker correleert ook hier aanzienlijk lager met de rapportcijfers dan de subtest Onzeker. Uit de tabel blijkt verder dat de samenhang tussen de cloze tests en het cijfer voor begrijpend lezen middelmatig tot aanzienlijk is. Opvallend is dat de MCCT bijna even hoog correleert met het cijfer voor technisch lezen als voor begrijpend lezen.

In tabel 5.29. worden de correlaties weergegeven tussen de scores van de leerlingen op de vijf ontwikkelde tests (inclusief de subtests voor Red.) die in de nameting werden afgenomen en de cijfers voor lezen, taal en rekenen van het Paasrapport. Opgemerkt moet worden dat de correlaties zijn berekend met de scores van de leerlingen van de controlescholen (5, 6, 7, 8)

Tabel 5.28. Correlaties tussen de vijf tests voor begrijpend lezen (voormeting) en praktijkoordelen voor lezen, taal en rekenen op het Kerstrapport: medianen van de correlaties per klas (het aantal klassen tussen haakjes).

	Leescijfer		Taalcijfer			Rekencijfer		
	Technisch lezen (14)	Begrijpend lezen (16)	Schrift. taalgebr. (16)	Mond. taalgebr. (16)	Dictee (17)	Hoofd-rekenen (15)	Cijferen (15)	Redactie rekenen (10)
HGD	.43	.57	.44	.36	.45	.52	.41	.45
BED	.40	.56	.46	.40	.44	.39	.36	.47
RED	.32	.46	.35	.37	.35	.47	.57	.55
Onzeker	.32	.43	.36	.41	.27	.47	.51	.50
Zeker	.24	.33	.24	.17	.22	.31	.33	.31
FRCT	.49	.64	.50	.37	.50	.44	.38	.44
MCCT	.50	.54	.52	.42	.54	.48	.47	.52

Tabel 5.29. Correlaties tussen de vijf tests voor begrijpend lezen (nameting) en praktijkoordelen voor lezen, taal en rekenen op het Paasrapport: medianen van de correlaties per klas (het aantal klassen tussen haakjes).

	Leescijfer		Taalcijfer			Rekencijfer		
	Technisch lezen (14)	Begrijpend lezen (17)	Schrift. taalgebr. (15)	Mond. taalgebr. (15)	Dictee (17)	Hoofd-rekenen (17)	Cijferen (17)	Redactie rekenen (10)
HGD	.49	.63	.49	.49	.44	.39	.48	.52
BED	.37	.61	.51	.48	.43	.44	.47	.52
RED	.40	.63	.51	.45	.53	.48	.54	.52
Onzeker	.17	.53	.44	.42	.32	.47	.40	.48
Zeker	.42	.53	.47	.43	.42	.38	.47	.43
FRCT	.50	.60	.59	.44	.54	.49	.50	.55
MCCT	.45	.62	.54	.54	.54	.45	.51	.55

én van drie experimentele scholen (1, 3, 4).

Uit tabel 5.29. blijkt dat de samenhang tussen de drie taakgerichte tests (vorm B) en het cijfer voor begrijpend lezen aanzienlijk is en beduidend hoger dan de samenhang tussen deze tests en het cijfer voor technisch lezen. De mediane correlaties tussen de drie taakgerichte tests en het cijfer voor begrijpend lezen liggen ook hoger dan de correlaties tussen deze tests en de cijfers voor taal en rekenen. Opvallend is dat de Redeneertest (vorm B) in de nameting hoger met het cijfer voor begrijpend lezen correleert dan in de voormeting; de correlatie met begrijpend lezen is nu hoger dan de correlatie met rekenen. De subtest Zeker correleert nu beduidend hoger met de betreffende rapportcijfers dan in de voormeting. Uit de tabel blijkt verder dat de samenhang tussen de cloze tests en het cijfer voor begrijpend lezen aanzienlijk is. De MCCT correleert nu beduidend hoger met het cijfer voor begrijpend lezen dan met het cijfer voor technisch lezen.

Het gevondene kunnen we als volgt samenvatten. De scores op de drie taakgerichte tests correleren beduidend hoger met het cijfer voor begrijpend lezen dan met het cijfer voor technisch lezen. De cloze tests correleren over het algemeen ook duidelijk hoger met het cijfer voor begrijpend lezen dan met het cijfer voor technisch lezen. De samenhang tussen de vijf ontwikkelde tests en het cijfer voor begrijpend lezen is over het algemeen hoger dan de samenhang tussen deze tests en de cijfers voor taal en rekenen.

Variantiecomponenten binnen de vijf tests

In deze sectie zal worden nagegaan in hoeverre er ruimte bestaat voor de veronderstelling dat de taakgerichte tests een specifiek aspect van begrijpend lezen meten. In een eerste analyse wordt vastgesteld welk percentage van de variantie in de ontwikkelde tests verklaard wordt door gestandaardiseerde tests voor technisch lezen, woordenschat en begrijpend lezen. Daarnaast wordt de omvang van de verschillende variantiecomponenten berekend. In een tweede analyse wordt vastgesteld welk percentage van de variantie in elke taakgerichte test bovendien nog wordt verklaard door de cloze tests en de beide andere taakgerichte tests. Bij deze analyses zal steeds een vergelijking worden gemaakt tussen het percentage niet verklaarde variantie en het percentage foutenvari-

antie om na te gaan in hoeverre er ruimte blijft voor een specifiek aspect van de taakgerichte tests.

•

Eerste analyse

Met behulp van regressie-analyse wordt voor elk van de vijf ontwikkelde tests (de afhankelijke variabelen) vastgesteld welk percentage variantie samenhangt met vier onafhankelijke variabelen en wat de omvang is van de verschillende variantiecomponenten. De vier onafhankelijke variabelen die in aanmerking komen, worden in drie blokken ingedeeld. Voor het eerste en het tweede blok is één test beschikbaar; het derde blok omvat meerdere tests. Het eerste blok duiden we aan met "technisch lezen" (TL). De test voor dit blok is de Eén-Minuut-Test van Brus en Voeten (1973). Het tweede blok wordt "woordenschat" (WS) genoemd. De test voor dit blok is de Woordenschattest van Stijnen (1975). Het derde blok duiden we aan met "begrijpend lezen" (BL). Voor dit blok zijn twee gestandaardiseerde tests beschikbaar nl. de test Schriftelijke Opdrachten van Brus en Van Bergen (1973) en de Stilleestest van Bakker (1972). Zoals reeds eerder vermeld, meten beide tests het begrijpend lezen in het algemeen. Vooruitlopend op de tweede analyse worden in één geval aan deze twee tests nog twee niet gestandaardiseerde tests toegevoegd nl. de Free Response Cloze Test en de Multiple Choice Cloze Test. Het spreekt voor zichzelf dat bij die gelegenheid de beide laatst genoemde tests niet als afhankelijke variabelen kunnen worden gebruikt.

Van de vijf afhankelijke variabelen zijn gegevens van drie meetmomenten bekend; van de vier onafhankelijke variabelen zijn de gegevens van één afname voorhanden. Voor de berekening van de variantiecomponenten maakt dit geen verschil. Voor elk meetmoment kunnen op dezelfde wijze de proporties variantie worden berekend. Voor de interpretatie maakt het echter wel verschil. Het is nl. bekend dat naarmate de tijdsduur tussen twee afnames toeneemt de correlatie afneemt. De tabellen die in deze sectie betrekking hebben op de voormeting, nameting en retentiemeting zijn om die reden niet geheel vergelijkbaar.

Een andere opsplitsing van de geobserveerde testvariantie is die in ware en foutenvariantie. Uitgaande van het klassieke testtheoretische model is deze opsplitsing te schrijven als $V_x = V_w + V_e$ of in proporties uitgedrukt:

$1 = V_w/V_x + V_e/V_x$. De term V_w/V_x wordt de betrouwbaarheid van de test genoemd en meestal aangeduid met r_{xx} . De proportie foutenvariantie van een test kan dus uitgedrukt worden als: $V_e/V_x = (1 - r_{xx})$. Deze proportie foutenvariantie wordt in de tabellen van deze sectie als een percentage aangegeven. Voor zover kan worden aangenomen dat de meetfouten van de verschillende tests onderling ongecorreleerd zijn en deze tevens ongecorreleerd zijn met de scores op de andere tests, is het mogelijk het percentage niet verklaarde variantie en het percentage foutenvariantie met elkaar te vergelijken. Uit de vergelijking van deze beide percentages kan men een indruk krijgen omtrent de mate waarin het percentage niet verklaarde variantie kan bestaan uit foutenvariantie. Dat deel van de niet verklaarde variantie dat niet als foutenvariantie kan worden aangemerkt, is een deel van de ware variantie dat niet samenhangt met de onafhankelijke variabelen.

Als maat voor de betrouwbaarheid wordt voor de ontwikkelde tests de KR-20 gekozen; bij de Multiple Choice Cloze Test is de alpha-coëfficiënt berekend. De betrouwbaarheidscoëfficiënten worden steeds op dat deel van de steekproef berekend waarop ook de regressievergelijkingen zijn berekend.

Alvorens de resultaten van de eerste analyse te presenteren, geven we eerst nog enige informatie omtrent de methode die gevolgd is bij de ontbrekende scores. Het subprogramma REGRESSION van de SPSS-pakket (Nie et al., 1975) vermeldt twee manieren om ontbrekende scores te behandelen nl. het paarsgewijs en het lijstgewijs weglaten van leerlingen die één of meer ontbrekende scores hebben. Beide methoden hebben nadelen. Het lijstgewijs weglaten kan resulteren in een sterke daling van het aantal leerlingen dat in de berekening wordt betrokken. Het paarsgewijs weglaten van leerlingen kan bij de berekening van correlatiecoëfficiënten resulteren in niet interpreteerbare statistische grootheden. Tegen het gebruik van de methode van paarsgewijs weglaten wordt in de SPSS-handleiding uitdrukkelijk gewaarschuwd.

Voor de berekening van de correlaties voor de hele steekproef zijn bij paarsgewijs weglaten ongeveer 500 leerlingen per variabele beschikbaar. Van hen hebben in totaal 383 leerlingen alle tests gemaakt. Gezien de omvang van de steekproef zou hier lijstgewijze weglating mogelijk zijn. Dit ligt echter anders als we ons beperken tot de drie controlescholen. Lijstgewijze weglating zou daar resulteren in kleine leerlingenaantallen. Om na te gaan

hoe ernstig de gevolgen zijn van paarsgewijze weglating, zijn op de resultaten van de voormeting beide weglatingmethoden toegepast. De verschillen in multiple R-kwadraat waren maximaal 3%, terwijl de verschillen in opgesplitste percentages kleiner waren. Bij beide methoden kwamen op dezelfde plaatsen zeer kleine negatieve percentages variantie voor. Zoals al eerder vermeld, mogen we er van uit gaan dat geen systematische factoren een rol hebben gespeeld bij het missen van scores. Omdat de verschillen gering zijn en het aantal leerlingen bij lijstgewijze weglating te klein wordt bij de controlescholen, is hier voor de paarsgewijze weglating van leerlingen gekozen bij de berekening van correlatiecoëfficiënten. De matrices met correlaties zijn in Bijlage 5.2. opgenomen. In deze matrices is ook het aantal leerlingen vermeld dat een betreffende test heeft gemaakt (N maximaal).

De resultaten van de regressie-analyses in de vorm van gekwadrateerde multiple correlatiecoëfficiënten en de decompositie van deze coëfficiënten in percentages gemeenschappelijke variantie zijn opgenomen in de tabellen 5.30. tot en met 5.34. Tabel 5.30. heeft betrekking op de gegevens van alle scholen uit de voormeting. Dit geldt ook voor tabel 5.31. Daarin is echter blok 3 uit de regressie-analyse, "begrijpend lezen", uitgebreid met de beide cloze tests. De tabellen 5.32. tot en met 5.34. hebben betrekking op drie controlescholen. De getallen in de tabellen zijn percentages variantie. De kleine verschillen die soms bestaan tussen het totale percentage verklaarde variantie en de som van de zeven componenten zijn het gevolg van afrondingsfouten.

De wijze waarop de variabelen binnen een blok samenhangen met de afhankelijke variabelen kan nagegaan worden door telkens de omvang van vier componenten te beoordelen nl. de eigen variantie van dat blok met de afhankelijke variabelen, de gemeenschappelijke variantie van dat blok en een van de beide andere blokken met de afhankelijke variabelen, en de gemeenschappelijke variantie van de drie blokken met de afhankelijke variabelen.

Blok 1: Technisch lezen.

De eigen variantie van de Eén-Minuut-Test met de afhankelijke variabelen is nagenoeg nihil. Uit de tabellen blijkt dat de maximale waarde 2% is en dit geldt dan nog alleen voor de Multiple Choice Cloze Test. Ook de gemeen-

schappelijke variantie van de Eén-Minuut-Test in combinatie met een van de beide andere blokken is nagenoeg nihil. De maximale waarde in alle vijf tabellen is 2%; dit geldt voor de Multiple Choice Cloze Test en de Free Response Cloze Test. In alle tabellen komen in deze kolommen een aantal (zeer kleine) negatieve getallen voor. Voor de interpretatie kunnen we deze getallen op nul stellen.

De conclusie kan luiden dat de variantie die de Eén-Minuut-Test gemeenschappelijk heeft met de vijf ontwikkelde tests ook gemeenschappelijk is met de variabelen in blok 2 en 3. Die keren dat sprake is van een zeer geringe eigen variantiecomponent of een variantiecomponent in combinatie met een van beide andere blokken is dit meestal het geval met een van beide of beide cloze tests. De omvang is echter zo gering dat we hier voorlopig geen betekenis aan willen toekennen. Verder onderzoek naar deze mogelijke samenhang zal daarom noodzakelijk zijn.

Blok 2: Woordenschat.

De eigen variantie van de Woordenschattest met de afhankelijke variabelen varieert rond de 5%. Maximaal komt 10% voor nl. met de Hoofdgedachtetest in de controlescholen, minimaal 1% met de subtest Zeker van de Redeneertest. Het percentage variantie dat van deze subtest verklaard wordt, is in de voormeting echter klein.

De gemeenschappelijke variantie van "woordenschat" en "begrijpend lezen" samen met de afhankelijke variabelen vormt steeds de grootste component van het opgesplitste percentage verklaarde variantie. In de meeste gevallen maakt deze component ongeveer de helft uit van de totale verklaarde variantie.

Blok 3: Begrijpend lezen.

Ook het percentage eigen variantie van blok 3 bedraagt ongeveer 5% met variaties in de tabellen van 1% voor de Hoofdgedachtetest in de voor- en nameting in de controlescholen tot 7 à 8% voor de Free Response Cloze Test en de Bedoelingtest in de controlescholen. De rest van de variantie van de afhankelijke variabelen die door de tests voor begrijpend lezen verklaard wordt, hangt samen met blok 2 of is gemeenschappelijk met beide overige blokken, zoals hiervoor reeds is beschreven.

Tabel 5.30. Percentages variantie van de ontwikkelde tests. (vorm A; voormeting: alle scholen). N = ca. 500.

	verklaarde variantie	gemeenschappelijk met				eigen variantie			niet verklaarde variantie	% fouten- variantie	betrouw- baarheid
		TL+WS+BL	TL+WS	TL+BL	WS+BL	TL	WS	BL			
HGD	53	10	0	0	31	0	9	3	47	11	89
BED	55	15	0	1	28	0	5	7	45	17	83
RED	37	6	0	0	23	0	4	4	63	17	83
Onzeker	35	6	0	0	22	0	4	4	65	17	83
Zeker	13	2	0	0	8	0	1	2	87	20	80
FRCT	65	20	0	2	32	1	6	5	35	14	86
MCCT	56	20	1	2	22	2	6	3	44	15	85

Tabel 5.31. Percentages variantie van de taakgerichte tests. (vorm A; cloze tests binnen blok 3; voormeting: alle scholen). N = ca. 500.

	verklaarde variantie	gemeenschappelijk met				eigen variantie			niet verklaarde variantie	% fouten- variantie	betrouw- baarheid
		TL+WS+BL	TL+WS	TL+BL	WS+BL	TL	WS	BL			
HGD	54	10	0	0	35	0	5	4	46	11	89
BED	56	15	0	1	30	0	3	7	44	17	83
RED	37	6	0	0	24	0	3	4	63	17	83
Onzeker	36	6	0	0	23	0	2	4	64	17	83
Zeker	13	2	0	0	9	0	1	2	87	20	80

Tabel 5.32. Percentages variantie van de ontwikkelde tests. (vorm A; voormeting: controlescholen 6, 7, 8). N=ca.220.

	verklaarde variantie	gemeenschappelijk met				eigen variantie			niet verklaarde variantie	% fouten- variantie	betrouw- baarheid
		TL+WS+BL	TL+WS	TL+BL	WS+BL	TL	WS	BL			
HGD	46	8	0	0	27	0	10	1	54	11	89
BED	47	10	0	0	25	0	4	7	53	17	83
RED	35	6	0	0	21	0	3	5	65	16	84
Onzeker	33	5	0	0	21	0	4	3	67	18	82
Zeker	15	3	0	0	8	0	1	3	85	18	82
FRCT	62	18	0	2	29	1	7	4	38	15	85
MCCT	53	16	1	2	21	2	7	5	47	16	84

Tabel 5.33. Percentages variantie van de ontwikkelde tests. (vorm B; nameting: controlescholen 6, 7, 8) N = ca. 220.

	verklaarde variantie	gemeenschappelijk met				eigen variantie			niet verklaarde variantie	% fouten- variantie	betrouw- baarheid
		TL+WS+BL	TL+WS	TL+BL	WS+BL	TL	WS	BL			
HGD	45	13	0	1	21	1	8	1	55	10	90
BED	48	12	0	1	25	0	6	3	52	16	84
RED	47	10	0	0	27	0	6	4	53	13	87
Onzeker	42	6	0	0	27	1	4	5	58	17	83
Zeker	32	9	0	1	15	0	5	2	68	19	81
FRCT	60	16	0	1	31	0	4	8	40	16	84
MCCT	51	16	0	2	23	1	5	3	49	13	87

Tabel 5.34. Percentages variantie van de ontwikkelde tests. (vorm A; retentiemeting: controlescholen 6, 7,8).
N = ca. 220.

	verklaarde variantie	gemeenschappelijk met				eigen variantie			niet verklaarde variantie	% fouten- variantie	betrouw- baarheid
		TL+WS+BL	TL+WS	TL+BL	WS+BL	TL	WS	BL			
HGD	50	13	0	1	25	0	9	2	50	11	89
BED	47	13	0	1	22	0	4	5	53	13	87
RED	46	10	0	0	25	0	7	4	54	11	89
Onzeker	40	6	0	0	25	0	6	3	60	17	83
Zeker	34	9	0	1	15	0	5	4	66	16	84
FRCT	55	16	0	1	26	1	6	5	45	16	84
MCCT	48	16	1	2	18	2	3	5	52	10	90

In tabellen 5.30. en 5.32 tot en met 5.34. bestaat blok 3 uit de Schriftelijke Opdrachten en de Stilleestest. In tabel 5.31. zijn daar de beide cloze tests aan toegevoegd. Vergelijking van tabel 5.30. en 5.31. laat zien dat daarmee het totale percentage verklaarde variantie met niet meer dan 1% toeneemt. Verdere verschillen tussen de twee tabellen doen zich voor met betrekking tot twee kolommen. Bij beide kolommen is de Woordenschattest betrokken. De eigen variantie van de Woordenschattest wordt enkele procenten kleiner, terwijl het percentage variantie dat door "woordenschat" en "begrijpend lezen" samen wordt verklaard wat groter wordt. We zien dus dat door toevoeging van de betreffende variabelen aan blok 3 het percentage verklaarde variantie nauwelijks toeneemt, maar dat een deel van de eigen variantie van de Woordenschattest nu ook gemeenschappelijk is met het uitgebreide blok "begrijpend lezen". De verschillen zijn echter gering.

Van de vijf ontwikkelde tests waarop met behulp van de gestandaardiseerde tests de regressie is uitgerekend, blijkt dat van de beide cloze tests het grootste percentage variantie wordt verklaard nl. 65% voor de Free Response Cloze Test en 56% voor de Multiple Choice Cloze Test. Deze variantie wordt voor ongeveer de helft verklaard door de gemeenschappelijke variantie van "woordenschat" en "begrijpend lezen"; 20% ervan wordt verklaard door de drie blokken gezamenlijk, terwijl ook nog enkele procenten verklaard worden door de blokken 2 en 3 afzonderlijk. In vergelijking met de taakgerichte tests valt het relatief hoge percentage op dat door de drie blokken gezamenlijk wordt verklaard.

Van de taakgerichte tests vertoont het percentage verklaarde variantie en de opdeling daarvan bij de Hoofdgedachtetest en de Bedoelingtest grote overeenkomst. In de voormeting wordt ruim de helft van de variantie verklaard; in de nameting en de retentiemeting wat minder maar hier speelt het tijdsverschil tussen de afnames mogelijk een rol. De variantie van deze beide tests hangt voor het grootste deel samen met de blokken 2 en 3 gezamenlijk. De samenhang met "technisch lezen" wordt alleen bepaald door de mate waarin "technisch lezen" samenhangt met "woordenschat" en "begrijpend lezen". Zoals vermeld is deze samenhang met "technisch lezen" voor de taakgerichte tests geringer dan voor de cloze tests. De variantie van de Redeneertest in de voormeting (tabel 5.30.) wordt slechts voor 37% door de drie blokken verklaard. De opsplitsing van de subtest Onzeker lijkt veel op die

van de totale test. Van de subtest Zeker wordt slechts 13% van de variantie verklaard.

Uit de tabellen 5.33. en 5.34. blijkt tenslotte dat het patroon van opsplitsing in de nameting en de retentiemeting hetzelfde blijft. De percentages verklaarde variantie verschillen meestal wat. Deze percentages worden lager voor de cloze tests en de Hoofdgedachte- en de Bedoelingtest, maar nemen toe voor de Redeneertest. Deze verandering van samenhang hebben we ook al eerder geconstateerd.

De tweede opsplitsing van de totale variantie van de afhankelijke variabelen is die in betrouwbare- en foutenvariantie. De betrouwbaarheid van de tests op de verschillende afnamemomenten en in de verschillende subgroepen varieert van .80 tot .90. Het percentage foutenvariantie varieert derhalve van 20% tot 10%. Vergelijken we het percentage foutenvariantie met het percentage niet verklaarde variantie dan blijkt dat het percentage niet verklaarde variantie telkens hoger is dan het percentage foutenvariantie. We kunnen de resultaten van de regressie-analyse dus niet afdoen door te stellen dat hetgeen niet verklaard wordt, zal bestaan uit foutenvariantie. Blijkbaar meet elk van de vijf tests nog wat anders dan "technisch lezen", "woordenschat" of "begrijpend lezen" zoals dat met de vier gestandaardiseerde tests gemeten wordt.

Tweede analyse

Uit de voorgaande analyse blijkt dat ongeveer de helft van de variantie in de drie taakgerichte tests verklaard wordt door vier gestandaardiseerde tests. Vooruitlopend op de vraag in hoeverre de drie taakgerichte tests mogelijkwerwijs een specifiek aspect van begrijpend lezen meten, is in deze analyse ook het percentage variantie berekend dat door de vier gestandaardiseerde tests en de beide cloze tests wordt verklaard. Vergelijking van tabel 5.30. en 5.31. maakt duidelijk dat door toevoeging van de cloze tests het totale percentage verklaarde variantie niet noemenswaard toeneemt; een aanzienlijk deel van de variantie in de drie taakgerichte tests blijft onverklaard door de zes tests op het gebied van lezen en woordenschat.

In de tweede analyse wordt de bovengenoemde procedure voortgezet en nagegaan welk percentage van de variantie in elke taakgerichte test verklaard wordt wanneer aan de vier gestandaardiseerde tests en de beide cloze tests

nog twee taakgerichte tests worden toegevoegd. Het resultaat van deze analyse wordt in tabel 5.35. weergegeven. In deze tabel zijn de percentages variantie opgenomen die niet verklaard worden door de vier gestandaardiseerde tests, de gestandaardiseerde tests en de beide cloze tests, en de gestandaardiseerde tests met de cloze tests en twee taakgerichte tests. In de tabel wordt ook het percentage foutenvariantie van elke taakgerichte test vermeld.

Tabel 5.35. Percentages niet verklaarde variantie van de taakgerichte tests. (vorm A; voormeting: alle scholen). N = ca. 500.

	niet verklaarde variantie door 4 gestand. tests	niet verklaarde variantie door 4 gestand. tests en 2 cloze tests	niet verklaarde variantie door 4 gestand. tests, 2 cloze tests en 2 taakgerichte tests	% foutenvariantie
HGD	47	46	38	11
BED	45	44	38	17
RED	63	63	57	17

Uit deze tabel blijkt dat door toevoeging van twee taakgerichte tests het percentage verklaarde variantie in de drie taakgerichte tests met ongeveer 7% toeneemt. Het percentage niet verklaarde variantie bedraagt voor de Hoofdgedachtetest en de Bedoelingtest 38% en voor de Redeneertest 57%. Deze percentages blijven beduidend hoger dan het percentage foutenvariantie van elke test. Elke taakgerichte test meet blijkbaar nog iets anders dan hetgeen door de bovengenoemde acht tests wordt gemeten. Er blijft dus ruimte voor de veronderstelling dat elke taakgerichte test een specifiek aspect van begrijpend lezen meet. Voor een meer definitieve uitspraak is cross-validatie-onderzoek gewenst.

5.1.5. De verandering in de prestaties op de ontwikkelde tests

Inleiding

In deze paragraaf wordt een beschrijving gegeven van de verandering in de prestaties van de leerlingen over drie perioden nl. de eerste periode tussen voormeting en nameting, de tweede periode tussen nameting en retentiemeting en de totale periode tussen voormeting en retentiemeting. De verandering in de prestatie van een leerling op een test tijdens deze drie periodes is het verschil tussen de score op de voormeting en nameting ($N - V$), de nameting en retentiemeting ($R - N$), en de voormeting en retentiemeting ($R - V$). Deze verschillscores of veranderingsscores kunnen positief of negatief zijn. In het eerste geval is sprake van vooruitgang in de prestaties, in het tweede geval van achteruitgang.

In deze paragraaf wordt eerst voor elke test een beschrijving gegeven van de verandering tijdens de totale periode. Er worden drie maten voor verandering gegeven nl. verschillen tussen gemiddelden, verschillen tussen percentages goed beantwoorde items en verschillen in percentages leerlingen die een "mastery-score" bereiken. Vervolgens wordt ingegaan op veranderingen tijdens de deelperiodes (de eerste en tweede periode). Daarna wordt een beeld gegeven van de grootte van de individuele verschillen in verandering. Ten slotte wordt een bespreking gewijd aan de problematiek om de samenhang vast te stellen tussen individuele verschillen in verandering en variabelen als technisch lezen, woordenschat en dergelijke. Opgemerkt moet worden dat de hier gepresenteerde beschrijving van de verandering in de prestaties van de leerlingen slechts beperkt van aard is. Aan de beschrijving ligt geen model ten grondslag.

De gegevens hebben uitsluitend betrekking op de controlescholen 6, 7 en 8. Controleschool 5 is buiten beschouwing gelaten om over de drie meetmomenten de beschikking te hebben over gelijke of bijna gelijke groepen leerlingen. De experimentele scholen werden weggelaten omdat effecten van de experimentele programma's bij de hier besproken veranderingen storend kunnen werken.

De totale periode

Tabel 5.36. geeft voor elke test de gemiddelde verandering tijdens de drie periodes.

Tabel 5.36. Gemiddelden en standaarddeviaties van verschillcores over de eerste periode (N - V), de tweede periode (R - N) en de totale periode (R - V). Controlescholen 6, 7, 8. N = ca. 220.

		N - V				R - N				R - V			
	s(v)	\bar{x}	$\frac{\bar{x}}{s(v)}$	s	$\frac{s}{s(v)}$	\bar{x}	$\frac{\bar{x}}{s(v)}$	s	$\frac{s}{s(v)}$	\bar{x}	$\frac{\bar{x}}{s(v)}$	s	$\frac{s}{s(v)}$
HGD	7.1	3.3	0.5	5.1	0.7	1.0	0.1	4.2	0.6	4.2	0.6	5.0	0.7
BED	5.1	1.6	0.3	4.1	0.8	1.1	0.2	3.7	0.7	2.6	0.5	4.2	0.8
RED	7.1	7.1	1.0	6.0	0.8	1.2	0.2	5.3	0.7	8.2	1.2	6.4	0.9
Onzeker	3.9	2.9	0.7	3.5	0.9	0.5	0.1	3.0	0.8	3.4	0.9	3.6	0.9
Zeker	5.2	4.2	0.8	4.8	0.9	0.7	0.1	3.7	0.7	4.8	0.9	5.3	0.9
FRCT	6.8	3.0	0.4	4.8	0.7	0.7	0.1	4.7	0.7	3.9	0.6	4.3	0.6
MCCT	14.8	3.4	0.2	10.4	0.7	2.0	0.1	9.0	0.6	5.7	0.4	12.0	0.8

\bar{x} = gemiddelde verschillscore

s = standaarddeviatie van verschillcores

s(v) = standaarddeviatie tijdens de voormeting van de controlescholen 6, 7, 8.

Beschouwen we eerst de vooruitgang tijdens de totale periode (R - V). De leerlingen zijn zoals verwacht op alle tests gemiddeld vooruit gegaan. Hierbij moet worden opgemerkt dat effecten van herhaalde afnames van (parallelvormen van) de tests steeds in de vooruitgang op de tests zijn inbegrepen. De gemiddelde vooruitgang varieert van 2.6 punten op de Bedoelingtest tot 8.2 punten op de Redeneertest. Deze getallen zijn echter onderling niet vergelijkbaar. Om vergelijkingen tussen de gemiddelde verandering op de diverse tests zinvoller te maken, wordt voor elke test de gemiddelde verandering gerelateerd aan de standaarddeviatie van de test. Deze standaarddeviaties zijn berekend over de gegevens van de controlescholen 6, 7 en 8 tijdens de voorme-

ting. Uit tabel 5.36. blijkt dat de gemiddelde vooruitgang, uitgedrukt in standaarddeviaties, weinig verschilt tussen de tests. De Redeneertest vormt hierop een uitzondering. De vooruitgang op de Redeneertest is groter dan op de andere tests. Hij bedraagt ca. 1 standaarddeviatie tegen ca. 0.5 standaarddeviatie op de andere tests (variërend van 0.4 op de MCCT tot 0.6 op de Hoofdgedachtetest en de FRCT).

Er zijn diverse verklaringen denkbaar voor de grotere vooruitgang op de Redeneertest. De meest waarschijnlijke verklaring lijkt samen te hangen met het beginniveau van de leerlingen op de test. Het is aannemelijk te veronderstellen dat de vooruitgang op een test groter zal zijn naarmate het beginniveau lager ligt.

Tabel 5.37. geeft voor elke test het percentage goed beantwoorde items op de voormeting, nameting en retentiemeting. Tevens worden de verschilpercentages tussen de meetmomenten weergegeven. Deze kunnen, evenals verschillen tussen gemiddelden uitgedrukt in standaarddeviaties, opgevat worden als maat voor verandering.

Tabel 5.37. Percentage goed beantwoorde items op voormeting (V), nameting (N) en retentiemeting (R).
Controlescholen 6, 7, 8. N = ca. 220.

	V	N	R	N-V	R-N	R-V
HGD	53	64	67	11	3	14
BED	60	66	70	6	4	10
RED	47	64	67	17	3	20
Onzeker	41	58	61	17	3	20
Zeker	51	68	70	17	2	19
FRCT	63	69	70	6	1	7
MCCT	69	73	76	4	3	7

Uit tabel 5.37. blijkt, dat er een verband bestaat tussen beginniveau (voortestscore) en vooruitgang; hoe lager het beginniveau, hoe groter de vooruitgang. Bij de Redeneertest is het beginniveau het laagst en daar treedt ook de grootste vooruitgang op.

Naast verschillen tussen gemiddelden uitgedrukt in standaarddeviaties en verschillen in percentages goed beantwoorde items wordt hier nog een derde maat gepresenteerd voor de verandering in de prestaties van de leerlingen. Het betreft verschillen in het percentage leerlingen dat een "mastery-score" bereikt. Voor elke test werd het percentage leerlingen bepaald dat een "mastery-score" behaalde op respectievelijk de voormeting, nameting en retentiemeting. Een "mastery-score" werd gedefinieerd als een score van minimaal 80% goed beantwoorde items van een test. De gegevens staan in tabel 5.38. Het verschil tussen de voormeting en de retentiemeting in het percentage leerlingen dat een "mastery-score" bereikt, varieert van 18% op de FRCT tot 29% op de subtest Zeker. Dit verschil is voor de Redeneertest niet of niet veel hoger dan voor de andere tests. De eerder geconstateerde grotere gemiddelde vooruitgang op de Redeneertest komt voornamelijk tot stand door een forse vooruitgang van veel leerlingen die geen "mastery-score" bereiken. Dit is in overeenstemming met het eerder genoemde lage beginniveau op de Redeneertest als verklaring voor de grotere gemiddelde vooruitgang op deze test.

Tabel 5.38. Percentage leerlingen met een "mastery-score" (80% of hoger).
Controlescholen 6, 7, 8. N = ca. 220.

	V	N	R	N-V	R-N	R-V
HGD	18	37	37	19	0	19
BED	21	37	46	16	9	25
RED	7	26	32	19	6	25
Onzeker	9	26	31	17	5	22
Zeker	14	34	43	20	9	29
FRCT	11	26	29	15	3	18
MCCT	26	39	48	13	9	22

De deelperiodes

De verandering tijdens de totale periode is opgebouwd uit de veranderingen tijdens de beide deelperiodes. De veranderingen per test tijdens de beide deelperiodes zijn al weergegeven in de tabellen 5.36. tot en met 5.38. Op

alle tests en voor alle drie maten is de vooruitgang tijdens de eerste periode groter dan tijdens de tweede periode. Hieronder worden enkele mogelijke verklaringen voor dit verschijnsel besproken.

- a. testvormen. In de retentiemeting wordt opnieuw dezelfde vorm afgenomen als in de voormeting. Toch treedt de grootste vooruitgang juist op tussen de voormeting en de nameting, waarbij in de nameting parallele testvormen zijn afgenomen. De vorm van de tests kan dus geen verklaring bieden.
- b. lengte van de beide periodes. De lengte van de eerste periode varieert, afhankelijk van de test en de school, van 2 - 3½ maand. De lengte van de tweede periode is 2 maanden. De eerste periode is dus over het algemeen langer dan de tweede en het is mogelijk dat dit verschil in lengte meespeelt in het verschil in vooruitgang.
- c. "underachievement" tijdens de voormeting. Het is mogelijk dat de leerlingen tijdens de voormeting prestaties geleverd hebben die onder hun werkelijke niveau liggen. Zoals ook in par. 5.1.4. wordt gesuggereerd, kunnen de omstandigheden tijdens de voormeting minder gunstig zijn geweest. De motivatie van de leerlingen was mogelijk iets minder, omdat ze al een aantal tests gemaakt hadden op het moment dat de hier besproken tests afgenomen werden. Bovendien kan het begrijpen van de taak tijdens de voormeting (eerste afname van de tests) voor sommige leerlingen moeilijkheden opgeleverd hebben; moeilijkheden die tijdens de nameting verminderd of verdwenen waren. Dit laatste geldt waarschijnlijk voor de Redeneertest en de Hoofdgedachtetest. Uit het percentage goed beantwoorde items (vgl. tabel 5.37.) blijkt dat deze tests bij de voormeting moeilijk waren voor de leerlingen.
- d. plafond-effect. De scores op de nameting zouden zo hoog kunnen zijn, dat daarna, tijdens de tweede periode, nog maar weinig vooruitgang mogelijk is. Uit tabel 5.37. is duidelijk dat dit niet het geval is. Het percentage goed beantwoorde items op de nameting staat voor elke test toe, dat er tijdens de tweede periode een vooruitgang plaatsvindt die minstens zo groot is als tijdens de eerste periode.
- e. seizoen-effect. Uit onderzoek (vgl. Brus & Voeten, 1973) is bekend dat in de eerste helft van het schooljaar een grotere leerwinst wordt behaald dan in de tweede helft. Het is mogelijk dat ook in dit onderzoek sprake

is van een zeker seizoen-effect.

Samenvattend lijken "underachievement" tijdens de voormeting, verschil in lengte van de beide periodes en mogelijk een seizoen-effect de meest plausibele verklaringen voor het verschijnsel dat de vooruitgang in de eerste periode groter is dan in de tweede periode.

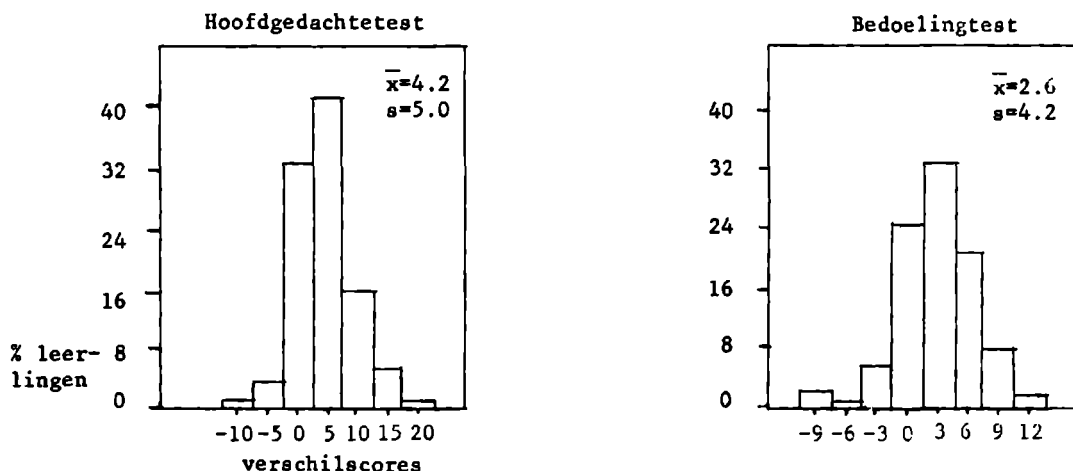
Individuele verschillen in verandering

We zullen nu een beeld geven van de grootte van de individuele verschillen in verandering. Eerst worden individuele verschillen in verandering tijdens de totale periode besproken, daarna tijdens de beide deelperiodes.

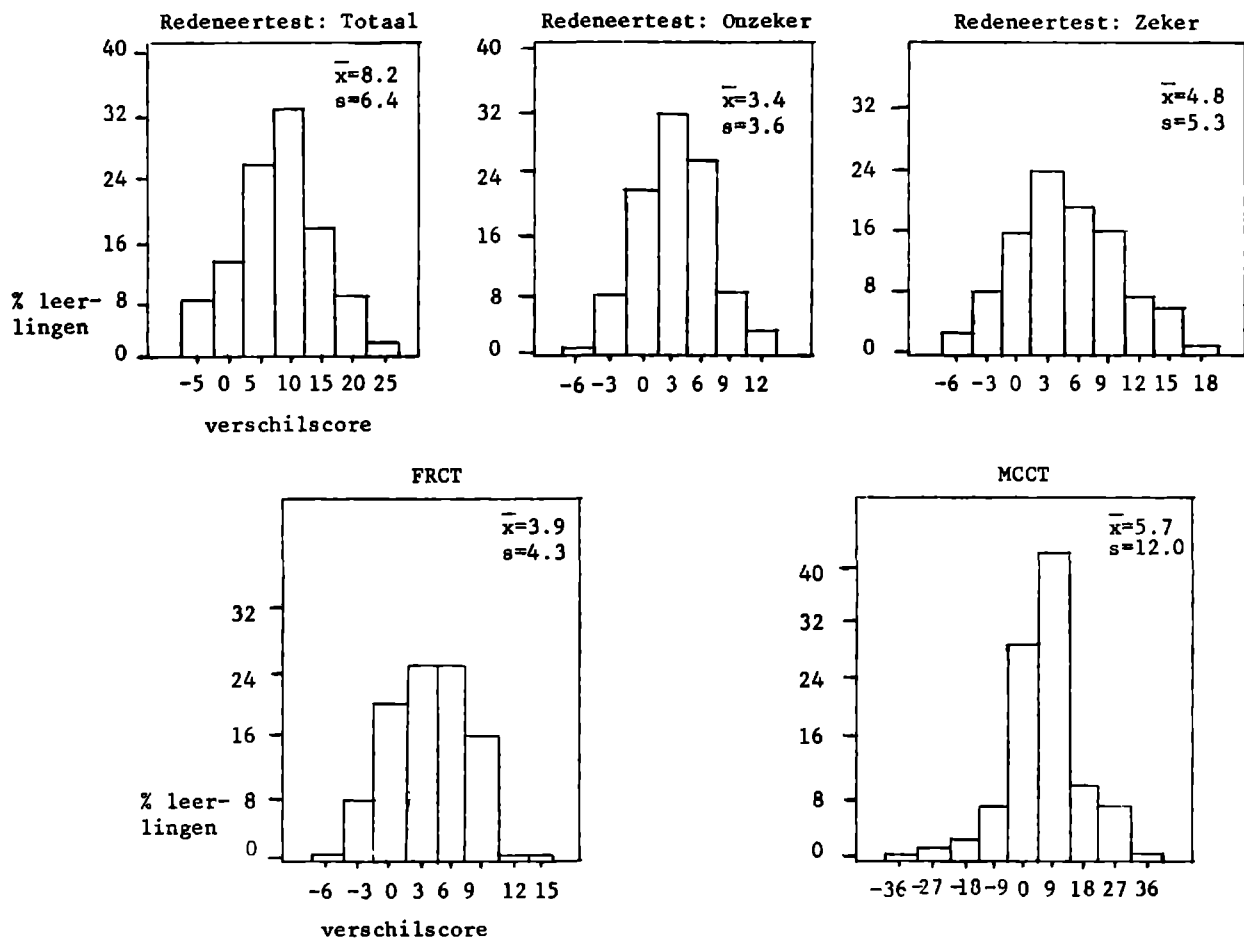
De totale periode

Voor elke leerling werd per test het verschil bepaald tussen de score op de retentiemeting en de score op de voormeting. De gemiddelden en standaarddeviaties van deze verschillen staan in tabel 5.36. Deze standaarddeviaties geven een indicatie van de grootte van de individuele verschillen in verandering. Een meer gedetailleerd beeld van de individuele verschillen in verandering geven de frequentieverdelingen van verschillen. Deze zijn weergegeven in figuur 5.10.

Figuur 5.10. Frequentieverdelingen van verschillen (retentiemeting-voormeting). Controlescholen 6, 7, 8. N = ca. 220.



Figuur 5.10. Vervolg.



Uit figuur 5.10. blijkt dat de leerlingen sterk verschillen in de verandering die ze vertonen tussen de voormeting en de retentiemeting. Bij elke test is een aantal leerlingen met een negatieve verschilscore; zij vertonen een achteruitgang in prestatie. Het percentage leerlingen met een negatieve verschilscore varieert van 11% op de Redeneertest tot 19% op de Bedoelingtest. Anderzijds laat figuur 5.10. ook zien dat er op elke test leerlingen zijn die zeer sterk vooruitgaan.

De individuele verschillen in verandering zijn niet op alle tests even groot. De standaarddeviaties van de verschillscores lopen nogal uiteen tussen de diverse tests (variërend van 3.6. op de Redeneersubtest Onzeker tot 12.0 op de MCCT). Analooq aan de gevolgde procedure bij de gemiddelde vooruitgang lijkt het ook hier raadzaam ter wille van de vergelijkbaarheid de standaarddeviaties van de verschillscores te relateren aan de standaarddeviaties van de tests. De laatste kolom van tabel 5.36. bevat de betreffende gegevens. De grootheid $\frac{s}{s(v)}$ varieert van 0.6 op de FRCT tot 0.9 op de Redeneertest en -subtests. Op deze wijze uitgedrukt variëren de individuele verschillen in verandering niet sterk tussen de tests; de grootste individuele verschillen in verandering treden op bij de Redeneertest.

De deelperiodes

Zijn de individuele verschillen in verandering tijdens de beide periodes even groot? Het antwoord op deze vraag is af te lezen uit tabel 5.36. De standaarddeviaties van verschillscores (al of niet gerelateerd aan de standaarddeviaties van de tests) zijn tijdens de eerste periode iets groter dan tijdens de tweede periode, maar de verschillen zijn gering. In tegenstelling tot de gemiddelde verandering zijn de individuele verschillen in verandering in beide periodes bijna even groot.

Factoren die samenhangen met individuele verschillen in verandering

Het onderzoek naar factoren die samenhangen met individuele verschillen in verandering wordt zeer gecompliceerd door de afhankelijkheid van verschillscores van het beginniveau van de leerlingen. De vooruitgang (verschilscore) zal over het algemeen groter zijn, naarmate het beginniveau lager is. Voor dit verschijnsel kunnen zowel een plafond-effect als een effect dat bekend staat als regressie naar het gemiddelde verantwoordelijk zijn. De afhankelijkheid van verschillscores van het beginniveau is af te lezen uit tabel 5.39. Hierin staan voor elke test de correlaties tussen de beginscores en de verschillscores over respectievelijk de eerste periode, de tweede periode en de totale periode. Voor de eerste en totale periode vormt de score op de voormeting de beginscore; over de tweede periode is de score op de nameting de beginscore.

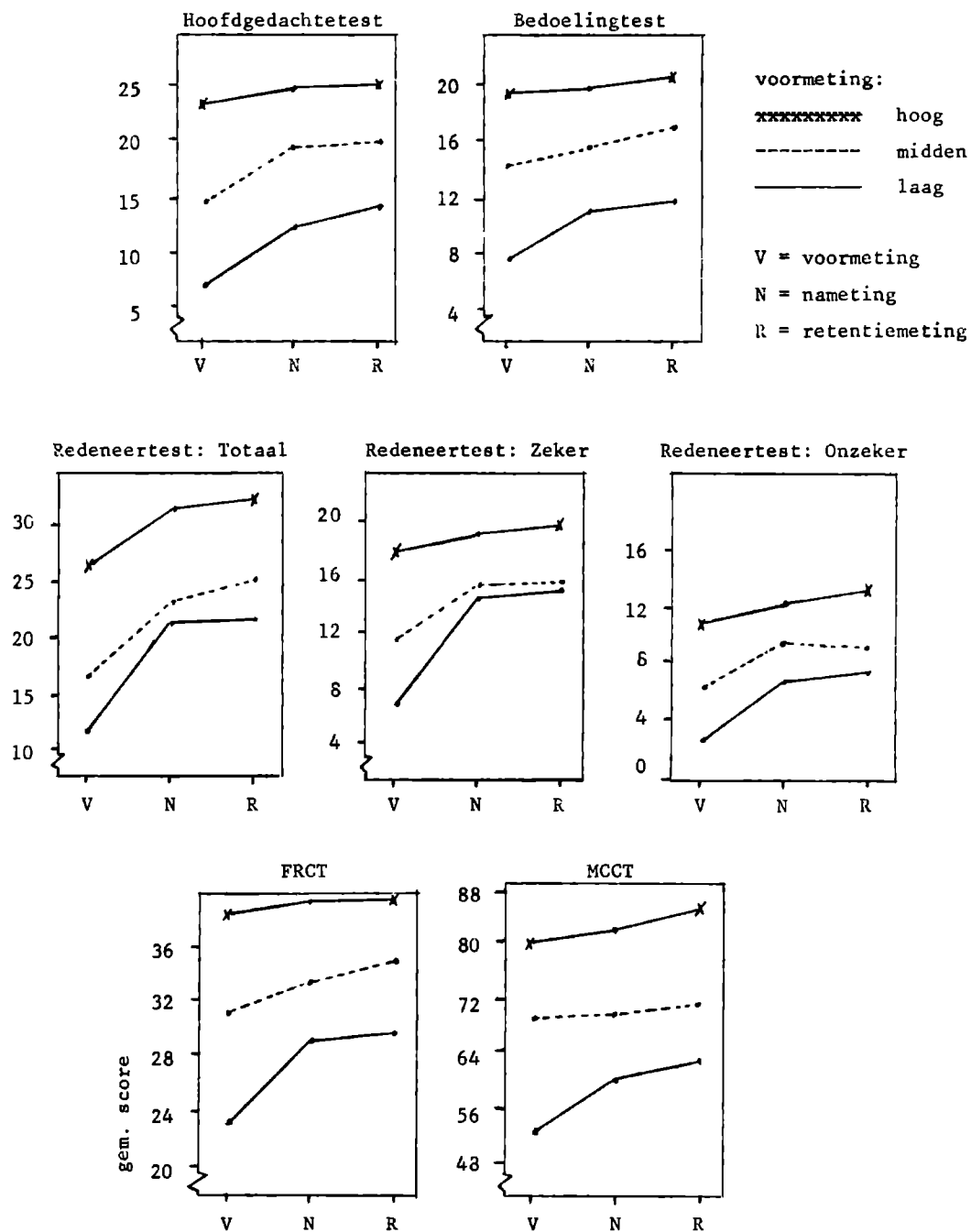
Tabel 5.39. Correlaties tussen beginscore en verschilscore over de eerste periode (N - V), de tweede periode (R - N) en de totale periode (R - V). Controlescholen 6, 7, 8. N = ca. 220.

	V x (N - V)	N x (R - N)	V x (R - V)
HGD	-.38	-.34	-.43
BED	-.41	-.31	-.37
RED	-.36	-.27	-.27
Onzeker	-.42	-.35	-.40
Zeker	-.58	-.32	-.54
FRCT	-.36	-.49	-.48
MCCT	-.34	-.09	-.28

Een consequentie van de afhankelijkheid van verschilscores van het begin-niveau is, dat verschilscores onderling niet direct vergelijkbaar zijn. Een-zelfde verschilscore zal moeilijker te behalen zijn naarmate het beginniveau van de verschilscore hoger ligt. Wanneer men de betreffende verschilscores correleert met andere variabelen wordt geïmpliceerd dat verschilscores wel onderling vergelijkbaar zijn, hetgeen resulteert in niet interpreteerbare correlatiecoëfficiënten.

Niettemin is het aantrekkelijk te onderzoeken hoe diverse variabelen (met name de Woordenschattest, Eén-Minuut-Test, Stilleestest, Standard Progressive Matrices en Schriftelijke Opdrachten) samenhangen met individuele verschillen in verandering. Overwogen is om partiële of semi-partiële correlaties te berekenen tussen verschilscores en scores op de Woordenschattest, de Eén-Minuut-Test en dergelijke. Uit deze overweging kon slechts één conclusie worden afgeleid nl. dat de betreffende correlaties weinig nieuwe informatie zouden opleveren. Uit een aantal berekeningen, die zijn uitgevoerd, bleek dit inderdaad het geval te zijn. De hoogste partiële of semi-partiële correlaties treden op bij die tests, die het hoogst correleren met de beginscores, te weten de Woordenschattest en Schriftelijke Opdrachten. Verdere pogingen om de samenhang vast te stellen tussen individuele verschillen in verandering en variabelen als woordenschat, technisch lezen en dergelijke, zijn - mede door het ontbreken

Figuur 5.11. Verandering in gemiddelde testscore over de meetmomenten per voormetingsniveau (laag, midden, hoog). Controlescholen 6, 7, 8. N = ca. 220.



van een model voor de beschrijving van veranderingen - niet ondernomen.

In aansluiting op tabel 5.39. wordt tot slot nog nadere informatie gegeven over het verband tussen beginniveau en verschilscore. Op elke test afzonderlijk werden de leerlingen op basis van de voormetingsscore uitputtend verdeeld in drie groepen van ongeveer gelijke grootte nl. een groep leerlingen met een lage voormetingsscore, een groep leerlingen met een middelmatige voormetingsscore en een groep leerlingen met een hoge voormetingsscore. Voor elk van deze drie groepen werden de gemiddelden en standaarddeviaties berekend op de voormeting, nameting en retentiemeting. Deze gegevens zijn verzameld in Bijlage 5.3. Figuur 5.11. geeft het verloop van de gemiddelde test-scores van de drie groepen over de meetmomenten.

Figuur 5.11. laat zien dat, zoals verwacht, de leerlingen met een lage voormetingsscore de grootste vooruitgang boeken. Verder valt op dat de drie groepen leerlingen steeds gescheiden blijven in die zin, dat de gemiddelde score op een later meetmoment van een lagere groep lager is dan de gemiddelde score van een hogere groep tijdens een vroeger meetmoment. Een uitzondering vormt de Redeneertest; hier heeft de lage groep op de nameting al een hogere gemiddelde score dan de middengroep op de voormeting.

Tenslotte is onderzocht hoe de veranderingen in testcores per school verlopen. Per school en per test werden gemiddelden en standaarddeviaties berekend op de voormeting, de nameting en de retentiemeting. De betreffende gegevens staan in Bijlage 5.4. De interpretatie van de gegevens wordt bemoeilijkt doordat de scholen een verschillende gemiddelde voormetingsscore hebben. De vooruitgang van school 7 lijkt voor de meeste tests iets groter dan van de beide andere scholen. Er zijn echter geen grote en consistente verschillen in verandering tussen de scholen.

5.1.6. Samenvatting

De vraagstellingen van het descriptieve gedeelte van het onderzoek hebben betrekking op vier onderwerpen nl. de prestaties van leerlingen van het vierde leerjaar op de ontwikkelde tests, de betrouwbaarheid van de tests, de validiteit van de tests en de verandering in prestaties over drie meetmomenten.

Wat betreft de prestaties van de leerlingen op de ontwikkelde tests werd eerst nagegaan in hoeverre de prestaties van de leerlingen van de betreffende onderzoeksscholen verschillen van de prestaties van een aselechte

steekproef van scholen. Door de betreffende onderzoeksscholen op een aantal variabelen te vergelijken met een steekproef die "representatief" kan worden geacht voor de landelijke populatie scholen, kon worden vastgesteld dat de prestaties van de leerlingen van de onderzoeksscholen waarschijnlijk hoger liggen dan die van de landelijke populatie scholen voor gewoon lager onderwijs.

De prestaties van de leerlingen op de vijf ontwikkelde tests werden vervolgens weergegeven in descriptieve maten en frequentieverdelingen. De frequentieverdelingen van de testscores wijken, met uitzondering van de Free Response Cloze Test, alle af van de normaalverdeling. De frequentieverdelingen zijn bevredigend van vorm. De enige test die een duidelijk positief scheve verdeling vertoont, is de Redeneertest.

Voor elke test werd een variantie-analyse uitgevoerd om na te gaan of de gemiddelde testcores van de onderscheiden sociale milieus van elkaar verschillen. Uit deze analyses bleek dat bij alle vijf tests de gemiddelde scores van de onderscheiden sociale milieus zeer significant verschillen. De percentages variantie die verklaard worden door de verschillen in sociale milieus variëren per test van maximaal 11% voor de Hoofdgedachtetest tot minimaal 7% voor de Multiple Choice Cloze Test. Bij alle tests verschillen ook de gemiddelde scores van de scholen significant van elkaar. De percentages variantie die verklaard worden door de verschillen tussen de scholen variëren per test van maximaal 7% voor de Hoofdgedachtetest en de Multiple Choice Cloze Test tot minimaal 3% voor de Redeneertest en de Free Response Cloze Test. Omdat in dit onderzoek slechts enkele gegevens zijn verzameld omtrent het sociaal milieu en de scholen is geen poging ondernomen om de invloeden van beide factoren op de testprestaties te analyseren, rekening houdend met de onderlinge afhankelijkheid van deze factoren.

De opzet van het onderzoek maakte het mogelijk om de vier hoofdmethoden die Drenth (1972) onderscheidt voor de bepaling van de betrouwbaarheid van tests toe te passen. De verschillende maten voor de betrouwbaarheid zijn berekend voor drie verschillende tijdstippen (voormeting, nameting, retentieme-

ting) en voor vier groepen van scholen. De Free Response Cloze Test heeft bij de voormeting een KR-20 van .86. Deze enigszins geflatteerde coëfficiënt blijft over de meetmomenten heen vrijwel gelijk met een lichte tendens tot dalen. De coëfficiënten van stabiliteit en a-specificiteit zijn bij deze test vrij hoog. De Multiple Choice Cloze Test heeft bij de voormeting een alpha-coëfficiënt van .85. Deze coëfficiënt stijgt over de meetmomenten naar .90. Uit de absolute hoogte van de correlaties tussen de meetmomenten blijkt dat deze test vrij stabiel is. De correlatie tussen de nameting en retentiemeting is bij deze test en de taakgerichte tests hoger dan de correlaties tussen de voormeting en nameting en de voormeting en retentiemeting. Het is waarschijnlijk dat tijdens de voormeting een aantal leerlingen prestaties heeft geleverd die beneden hun eigenlijke niveau liggen. De Hoofdgedachtetest heeft bij de voormeting een KR-20 van .89. De KR-20 blijft voor alle groepen scholen vrijwel constant over de meetmomenten. Uit de absolute hoogte van de correlaties tussen de meetmomenten blijkt dat deze test vrij stabiel is. De Bedoelingtest heeft bij de voormeting een KR-20 van .83. De KR-20 blijft voor alle groepen scholen vrijwel constant over de meetmomenten. Uit de absolute hoogte van de correlaties tussen de meetmomenten blijkt dat deze test vrij stabiel is. De grootste stabiliteit treedt op tussen de nameting en retentiemeting. De Redeneertest heeft bij de voormeting een KR-20 van .83. De KR-20 vertoont voor alle groepen scholen een stijging over de meetmomenten, die waarschijnlijk voor een groot deel het gevolg is van een grotere spreiding. De test is vrij stabiel bij de experimentele scholen. De subtest Zeker heeft bij de voormeting een KR-20 van .80 en neemt over de meetmomenten iets toe. De correlaties tussen de voormeting en nameting en de voormeting en retentiemeting zijn laag. De subtest Onzeker heeft bij de voormeting een KR-20 van .83. De KR-20 blijft over de meetmomenten vrijwel constant. Deze subtest is evenals de subtest Zeker vrij instabiel tussen voormeting en nameting.

Uit de hoogte van de coëfficiënten van stabiliteit, a-specificiteit en interne consistentie kan worden afgeleid dat de vijf ontwikkelde tests voldoen aan de eisen van betrouwbaarheid.

In verband met het onderzoek naar de (begrips)validiteit van de vijf ontwikkelde tests werden verschillende analyses uitgevoerd. Om op een enigszins andere wijze een beeld te krijgen van de homogeniteit van de tests werden factoranalyses uitgevoerd op de items c.q. subtests van de verschillende tests. Uit de resultaten van deze analyses blijkt dat de Hoofdgedachtetest, de Bedoelingtest en de Multiple Choice Cloze Test één factor meten. De Redeneertest bevat twee interpreteerbare factoren. Over de homogeniteit van de Free Response Cloze Test kan met behulp van factoranalyse waarschijnlijk geen zinvolle uitspraak worden gedaan.

Op basis van de veronderstelling dat de vijf tests in sterke mate een beroep doen op hogere cognitieve processen werd verwacht dat de tests in aanzienlijke mate zouden samenhangen. Wat betreft de samenhang tussen de drie taakgerichte tests wordt deze verwachting grotendeels bevestigd met name bij de nameting en retentiemeting. De samenhang tussen de drie taakgerichte tests en de cloze tests is bij de voormeting middelmatig. Bij de nameting en retentiemeting is de samenhang tussen de taakgerichte tests en de Multiple Choice Cloze Test aanzienlijk tot hoog, terwijl de samenhang tussen de taakgerichte tests en de Free Response Cloze Test varieert van middelmatig tot aanzienlijk. De Multiple Choice Cloze Test blijkt overeenkomstig de verwachting geen maat te zijn voor een specifiek aspect van begrijpend lezen i.c. letterlijk begrijpend lezen.

Volgens verwachting correleren de drie taakgerichte tests in het vierde leerjaar laag en de cloze tests middelmatig met een maat voor technisch lezen i.c. de Eén-Minuut-Test van Brus en Voeten (1973). Eveneens volgens de verwachting correleren de vijf ontwikkelde tests over het algemeen in aanzienlijke mate met een maat voor begrijpend lezen i.c. de test Schriftelijke Opdrachten van Brus en Van Bergen (1973). De ontwikkelde tests vertonen tegen de verwachting in slechts een middelmatige samenhang met een andere maat voor begrijpend lezen i.c. de Stilleestest van Bakker (1972). Een verklaring voor deze middelmatige samenhang zou gevonden kunnen worden in het gegeven dat de Stilleestest niet voldoende betrouwbaar is. De drie taakgerichte tests vertonen in het algemeen een aanzienlijke samenhang met een maat voor leeswoordenschat i.c. de Woordenschattest van Stijnen (1975). De samenhang tussen de cloze tests en de betreffende Woordenschattest varieert van aanzienlijk tot hoog. Deze resultaten komen grotendeels overeen met de eerder

geformuleerde verwachting. Overeenkomstig de verwachting vertonen de drie taakgerichte tests en de cloze tests slechts een middelmatige samenhang met een maat voor non-verbale intelligentie i.c. de Standard Progressive Matrices van Raven (1958).

Wat betreft de samenhang tussen de scores op de vijf ontwikkelde tests en de rapportcijfers voor lezen, taal en rekenen werd het volgende geconstateerd. De scores op de drie taakgerichte tests correleren beduidend hoger met het cijfer voor begrijpend lezen dan met het cijfer voor technisch lezen. De cloze tests correleren over het algemeen ook duidelijk hoger met het cijfer voor begrijpend lezen dan met het cijfer voor technisch lezen. De samenhang tussen de vijf ontwikkelde tests en het cijfer voor begrijpend lezen is over het algemeen hoger dan de samenhang tussen deze tests en de cijfers voor taal en rekenen.

Met behulp van multiple regressie-analyse is voor elk van de vijf ontwikkelde tests het percentage variantie berekend dat elk van deze tests gemeenschappelijk heeft met drie (blokken van) variabelen nl. technisch lezen (gemeten met de Eén-Minut-Test), woordenschat (gemeten met de Woordenschattest) en begrijpend lezen (gemeten met een combinatie van tests nl. de test Schriftelijke Opdrachten en de Stilleestest). Het gemeenschappelijke percentage variantie van elk van de tests met elk blok van onafhankelijke variabelen is gering nl. ongeveer 5% voor woordenschat en voor begrijpend lezen en nagenoeg nihil voor technisch lezen. De blokken "woordenschat" en "begrijpend lezen" samen verklaren ongeveer de helft van de totale verklaarde variantie. Het blok "technisch lezen" draagt afzonderlijk of in combinatie met een van de beide andere blokken nagenoeg niets bij aan de verklaring van de variantie. Van de cloze tests wordt de meeste variantie verklaard; maximaal 65% van de Free Response Cloze Test. Van de taakgerichte tests wordt wat minder variantie verklaard. Van de Redeneertest wordt tijdens de voormeting maximaal 37% verklaard; bij de nameting en retentiemeting echter 10% meer. Het patroon van opsplitsingen is voor alle tests in grote lijnen hetzelfde: 50% van de verklaarde variantie wordt verklaard door "woordenschat" en "begrijpend lezen" gezamenlijk, een klein deel door deze twee blokken afzonderlijk en de rest door de drie blokken gezamenlijk. Vergelijking van de percentages niet verklaarde variantie met de percentages foutenvariantie toont aan, dat hetgeen niet verklaard kan worden door de gestandaardiseerde tests niet al-

leen toegeschreven kan worden aan mogelijke meetfouten. Het percentage verklaarde variantie in elke taakgerichte test neemt ongeveer 7% toe wanneer aan de vier gestandaardiseerde tests de beide cloze tests en twee taakgerichte tests worden toegevoegd. De percentages niet verklaarde variantie blijven ook in dit geval beduidend hoger dan het percentage foutenvariantie van elke test. Hieruit kan worden afgeleid dat er ruimte blijft voor de veronderstelling dat elke taakgerichte test een specifiek aspect van begrijpend lezen meet.

Met betrekking tot de validiteit van de vijf ontwikkelde tests kan in het algemeen worden geconcludeerd dat de resultaten grotendeels overeenkomen met de verwachtingen. Een definitief antwoord op de vraag naar de validiteit van de tests kan in dit verband echter niet worden gegeven. Verder onderzoek is hiervoor noodzakelijk.

Wat betreft de verandering in de prestaties op de ontwikkelde tests werden voor de totale periode (voormeting - retentiemeting) en de beide deelperiodes (voormeting - nameting en nameting - retentiemeting) drie maten gepresenteerd nl. verschillen tussen gemiddelden, verschillen tussen percentages goed beantwoorde items en verschillen in percentages leerlingen die een "mastery-score" behalen. Uit de gegevens blijkt dat over de totale periode de gemiddelde vooruitgang, uitgedrukt in standaarddeviaties, weinig verschilt tussen de tests en ca. 0.5 standaarddeviatie bedraagt. De Redeneertest vormt hierop een uitzondering; bij deze test bedraagt de gemiddelde vooruitgang ca. 1 s. De meest waarschijnlijke verklaring hiervoor lijkt te liggen in het lagere beginniveau (voormetingsscore) van de leerlingen op de Redeneertest. Uit de gegevens blijkt verder dat op alle tests de vooruitgang tijdens de eerste periode (voormeting - nameting) groter is dan tijdens de tweede periode (nameting - retentiemeting). De meest plausibele verklaringen voor dit verschijnsel lijken "underachievement" tijdens de voormeting, verschil in lengte van de beide periodes en mogelijk een seizoen-effect.

Wat betreft de individuele verschillen in verandering tijdens de drie periodes valt op dat de leerlingen sterk verschillen tussen de voormeting en de retentiemeting. Het percentage leerlingen met een negatieve verschilscore varieert in deze periode van 11% op de Redeneertest tot 19% op de Bedoelingstest. De meeste leerlingen gaan echter vooruit, sommigen zelfs zeer sterk.

De individuele verschillen in verandering, uitgedrukt in standaarddeviaties van de tests, variëren in deze periode niet sterk tussen de tests. In tegenstelling tot de gemiddelde verandering zijn de individuele verschillen in verandering in de beide deelperiodes bijna even groot.

Het onderzoek naar factoren die samenhangen met individuele verschillen in verandering werd zeer bemoeilijkt door de afhankelijkheid van de verschillen van het beginniveau van de leerlingen. Door deze afhankelijkheid zijn de verschillen onderling niet direct vergelijkbaar en zijn de correlaties van deze verschillen met andere variabelen zoals technisch lezen, woordenschat, non-verbale intelligentie en dergelijke niet interpreteerbaar. Andere mogelijke bewerkingen zoals partiële of semi-partiële correlaties leveren weinig nieuwe informatie op. Tenslotte zijn de leerlingen - bedoeld worden de leerlingen van drie controlescholen - op elke test afzonderlijk verdeeld in drie groepen op basis van de voormetingsscore nl. een groep leerlingen met een lage voormetingsscore, een groep met een middelmatige voormetingsscore en een groep met een hoge voormetingsscore. Uit het verloop van de prestaties over de drie meetmomenten blijkt opnieuw de afhankelijkheid van de verschillen van het beginniveau van de leerlingen. De leerlingen met een lage voormetingsscore gaan relatief het sterkst vooruit.

5.2. Het experimentele gedeelte van het onderzoek

5.2.1. Hypothesen

In hoofdstuk IV zijn drie vraagstellingen van het experimentele gedeelte van het onderzoek geformuleerd. De eerste vraag heeft betrekking op het effect op korte en lange termijn van elk van de drie taakgerichte programma's op de beheersing van de desbetreffende taken, gemeten met de daarmee corresponderende taakgerichte tests. De tweede vraag heeft betrekking op het effect op korte en lange termijn van elk taakgericht programma op de beheersing van die taken waarop de beide andere programma's zijn gericht en die gemeten worden met daarmee corresponderende tests. De derde vraag heeft betrekking op het effect op korte en lange termijn van elk programma op de ontwikkeling van begrijpend lezen in het algemeen, gemeten met de Free Response Cloze Test en de Multiple Choice Cloze Test.

Het antwoord op de eerste vraag is didactisch van belang omdat hierdoor een indicatie kan worden verkregen van de waarde i.c. het effect van taakgericht onderwijs op het gebied van begrijpend lezen van teksten (de zogenaamde "skill-approach"). Verwacht wordt dat de taakgerichte programma's, die expliciet aandacht besteden aan de ontwikkeling van vooral strategische of procedurele kennis met betrekking tot de deeltaken en de eindtaak, een positief effect zullen hebben op de met de eindtaken corresponderende tests. Naar verwachting resulteren de taakgerichte programma's in een hoger prestatieniveau bij die groepen van leerlingen die met behulp van deze programma's onderwijs ontvangen in vergelijking met een controlegroep die een programma krijgt aangeboden dat niet op een specifieke leestaak is gericht.

De tweede vraag vormt een nadere concretisering van de vraag naar de realiteitswaarde van de onderscheiding van de drie geselecteerde taken op het gebied van begrijpend lezen. De vraag in hoeverre de betreffende taken empirisch onderscheidbaar zijn, wordt primair gesteld vanuit een didactische achtergrond nl. de legitimiteit van taakgericht onderwijs op het gebied van begrijpend lezen. Indien zou blijken dat de taken in werkelijkheid niet zijn te onderscheiden, vervalt een belangrijk motief voor taakgericht onderwijs. Zoals reeds eerder vermeld, verwachten we dat de drie geselecteerde taken verschillen en dat een programma dat op één van deze taken is gericht geen positief effect zal hebben op de beheersing van die taken waarop de beide andere programma's zijn gericht. Opgemerkt moet worden dat de vraag naar het transfer-effect van de programma's alleen kan worden beantwoord als de programma's zelf een positief effect hebben.

Wat betreft de derde vraag verwachten we niet dat één taakgericht programma een positief effect zal hebben op het begrijpend lezen in het algemeen, zoals gemeten met de beide cloze tests. Deze verwachting is gebaseerd op het idee dat een taakgericht programma slechts op een beperkt aspect van het begrijpend lezen van teksten is gericht en dat alleen een reeks van taakgerichte programma's invloed kan hebben op het begrijpend lezen in het algemeen. Ook hier moet worden opgemerkt dat de vraag naar een mogelijk transfer-effect van de programma's alleen kan worden beantwoord als de programma's zelf een positief effect hebben.

In hypothese-vorm kunnen de bovenstaande verwachtingen voor zowel de korte als de lange termijn als volgt worden geformuleerd:

- Hypothese 1 Een taakgericht programma op het gebied van het begrijpend lezen van teksten resulteert bij leerlingen die dit programma krijgen aangeboden in een hogere beheersing van de betreffende taak gemeten met een daarmee corresponderende test in vergelijking met leerlingen die een programma ontvangen dat niet op een specifieke leestaak is gericht.
- Hypothese 2 Een taakgericht programma op het gebied van het begrijpend lezen van teksten resulteert bij leerlingen die dit programma krijgen aangeboden niet in een hogere beheersing van twee andere taken gemeten met daarmee corresponderende tests in vergelijking met leerlingen die een programma ontvangen dat niet op een specifieke leestaak is gericht.
- Hypothese 3 Een taakgericht programma op het gebied van het begrijpend lezen van teksten resulteert bij leerlingen die dit programma krijgen aangeboden niet in een hogere beheersing van begrijpend lezen zoals gemeten met de cloze tests in vergelijking met leerlingen die een programma ontvangen dat niet op een specifieke leestaak is gericht.

Het controleprogramma wordt in de bovenstaande hypothesen aangeduid met een programma dat niet op een specifieke leestaak is gericht. De drie bovengenoemde hypothesen kunnen nog verder worden gespecificeerd door voor de taakgerichte programma's één van de drie experimentele programma's in te vullen en voor de taakgerichte tests één van de drie tests. Een opsomming van alle volledig gespecificeerde hypothesen wordt weinig zinvol geacht. Bij de toetsing van de hypothesen zal uiteraard wel rekening worden gehouden met de verdere uitsplitsing.

Zoals ook uit de opzet van het onderzoek blijkt, worden de hypothesen eerst getoetst in een analyse binnen de experimentele scholen. Daarna worden deze hypothesen nogmaals getoetst in een analyse waarbij zowel de experimentele scholen als de controlescholen betrokken zijn. In beide gevallen zal een analyse worden uitgevoerd voor zowel de nameting als de retentiemeting.

5.2.2. Data

Uit de opzet van het onderzoek blijkt dat er sprake is van twee groepen van scholen nl. experimentele scholen en controlescholen. Op de eerste groep

van scholen werden drie experimentele programma's en een controleprogramma aangeboden, terwijl op de andere groep geen speciale programma's werden gegeven. Binnen de experimentele scholen werden de leerlingen van de vierde klassen van elke school op basis van een factorscore uit de voormeting in vier vergelijkbare groepen ingedeeld. Elke groep ontving gedurende vijf weken onderwijs in een van de vier programma's nl. Hoofdgedachte, Bedoeling, Redeneren en het Controleprogramma. Het laatste programma was bestemd voor de controlegroep. Tijdens drie periodes werden op zowel de experimentele scholen als de controlescholen twee parallelle vormen van de vijf ontwikkelde tests afgenomen.

Volgens de bovenbeschreven opzet kunnen de gegevens van de experimentele scholen schematisch worden weergegeven als in figuur 5.12. Dit observatie-design is voor de analyse vereenvoudigd tot een design waarin nog twee gekruiste factoren resteren nl. "school" (Factor A) en "programma" (Factor B) en waarin de tests de afhankelijke variabelen vormen. De meetmomenten vormen geen aparte factor. Voor meetmoment M2 en M3 wordt afzonderlijk een analyse uitgevoerd.

Wat betreft de ontbrekende gegevens is reeds in hoofdstuk IV vermeld dat de gegevens van één experimentele school voor zowel de nameting als de retentiemeting buiten de analyses moesten worden gehouden, omdat getwijfeld werd aan de betrouwbaarheid van een deel van de gegevens. Eveneens is vermeld dat één controleschool niet aan de retentiemeting kon meewerken.

Ook binnen de resterende scholen is sprake van uitval omdat een (beperkt) aantal leerlingen een of meerdere tests tijdens de nameting of retentiemeting misten. Daar ontbrekende scores in de dataset onoverkomelijke problemen opleveren voor de geplande analyses, werd besloten de scores van leerlingen die tijdens de nameting, de retentiemeting of tijdens beide metingen samen twee of meer tests hadden gemist bij de analyses buiten beschouwing te laten. Hierdoor werden de scores van in totaal 5 leerlingen uit de experimentele scholen verwijderd. Van de 260 leerlingen uit de controlescholen werden om dezelfde reden de scores van 22 leerlingen buiten beschouwing gelaten. Ook werd het ongewenst geacht om leerlingen uit de experimentele scholen, die meer dan twee van de tien lessen hadden gemist in de analyses te betrekken. Hierdoor werden de scores van 4 leerlingen bij de analyses van de gegevens eveneens buiten beschouwing gelaten. Het aantal resterende leerlingen wordt

Figuur 5.12. Datamatrix voor de gegevens van de experimentele scholen.

		M1					M2					M3				
		T1	T2	T3	T4	T5	T1	T2	T3	T4	T5	T1	T2	T3	T4	T5
S1	P1															
	P2															
	P3															
	P4															
S2	P1															
	P2															
	P3															
	P4															
S3	P1															
	P2															
	P3															
	P4															

waarin: M = Meetmoment

M1 = voormeting: november-december 1978, afname A-vormen.

M2 = nameting: maart 1979, afname B-vormen.

M3 = retentiemeting: mei-juni 1979, afname A-vormen.

T = Tests

De afhankelijke variabelen. Dit zijn achtereenvolgens:

T1 = de Hoofdgedachtetest (HGD)

T2 = de Bedoelingtest (BED)

T3 = de Redeneertest (RED)

T4 = de Free Response Cloze Test (FRCT)

T5 = de Multiple Choice Cloze Test (MCCT)

S = Scholen

Van de vier experimentele scholen met 9 vierde klassen resteerden na uitval nog drie scholen (S1, S2, S3) met zeven klassen.

P = Programma's

P1 = programma Hoofdgedachte

P2 = programma Bedoeling

P3 = programma Redeneren

P4 = Controleprogramma

De leerlingen (niet aangegeven) zijn genesteld binnen elke rij van dit schema, d.w.z. binnen elke programmagroep per school.

Tabel 5.40. Aantal leerlingen per school en per experimentele conditie.

Experimentele scholen

	school 1	school 2	school 3	totaal
Programma Hoofdgedachte	16	18	12	46
Programma Bedoeling	15	19	10	44
Programma Redeneren	16	18	12	46
Controleprogramma	16	18	12	46
Totaal	63	73	46	182

Controlescholen

	school 4	school 5	school 6	school 7	totaal
nameting	57	79	75	49	260
retentiemeting	57	79	75	-	211

Tabel 5.41. Aantal geschatte scores.

	Experimentele scholen				Controlescholen	
	HGD progr.	BED progr.	RED progr.	Controleprogr.		
Hoofdgedachtetest			(1)	(1)	10	(5)
Bedoelingtest	(3)	2			7	(2)
Redeneertest	(5)	(1)		(1)	6	(4)
FRCT	(2)		2 (2)	2	5	(3)
MCCT	2 (1)	2	(1)		7	(3)

De getallen die niet tussen haakjes staan hebben betrekking op de nameting.

De getallen tussen haakjes hebben betrekking op de retentiemeting.

- uitgesplitst naar school en experimentele conditie - vermeld in tabel 5.40.

Onder het in tabel 5.40. genoemde aantal leerlingen bevinden zich ook leerlingen die óf bij de nameting óf bij de retentiemeting óf bij beide metingen één van de vijf tests hebben gemist. Wanneer deze leerlingen in de analyses ook buiten beschouwing worden gelaten, zou dit resulteren in een verdere daling van het aantal leerlingen dat beschikbaar is. Besloten werd voor deze leerlingen de ontbrekende score te schatten. Een ontbrekende score werd met behulp van de vier wel aanwezige testcores geschat in een regressievergelijking. Voor elke test werd zowel voor de nameting als voor de retentiemeting en binnen elk van de vier experimentele condities en binnen de groep van controlescholen een regressievergelijking berekend. In totaal werden op deze wijze 40 verschillende regressievergelijkingen berekend, die echter niet alle zijn gebruikt bij het schatten van de scores. Een overzicht van het aantal scores dat op deze wijze is geschat, wordt gegeven in tabel 5.41.

5.2.3. Analyse van de gegevens binnen de experimentele scholen

Analyseprocedure

De eerder geformuleerde hypothesen hebben alle betrekking op een vergelijking tussen de taakgerichte programma's en het controleprogramma. Deze hypothesen kunnen met behulp van t-toetsen worden getoetst. Hypothese 1 vraagt een éézijdige toetsing, terwijl voor hypothesen 2 en 3 een tweezijdige toetsing nodig is. Gekozen wordt voor een significantieniveau van 5% bij éézijdige toetsing en van 10% bij tweezijdige toetsing. Deze toetsingen worden uitgevoerd in het kader van één multivariate covariantie-analyse. Door middel van deze analyse kunnen de hypothesen worden getoetst op basis van alle afhankelijke variabelen gezamenlijk, waarbij tevens rekening wordt gehouden met de intercorrelaties tussen de afhankelijke variabelen. Naast het effect van programma's wordt in de analyse ook het effect van scholen betrokken om na te gaan of een eventueel effect van de programma's gegeneraliseerd kan worden over de drie experimentele scholen. Als afhankelijke variabelen worden de drie taakgerichte tests en de beide cloze tests gebruikt. De analyse wordt afzonderlijk uitgevoerd voor de nameting en retentiemeting.

Als covariabele wordt de factorscore uit de voormeting gebruikt. Deze factorscore vormt per leerling een z-score, berekend uit de eerste principale component van tests die betrekking hebben op begrijpend lezen. Dit is dezelfde factorscore die vóór de aanvang van het experiment is gehanteerd om de leerlingen per school in te delen in vier vergelijkbare groepen (vgl. par. 4.8.6.). Het effect van de covariabele in de uitgevoerde analyse valt af te leiden uit de omvang van de reductie van de error-termen. Deze staan vermeld in Bijlage 5.5., tabel b. Van het gebruik van de scores op de vijf ontwikkelde tests uit de voormeting als covariabelen is afgezien, omdat te veel leerlingen op één of meer van de tests een score misten.

De factoren A en B worden beide opgevat als "fixed". Voor Factor B (programma's) ligt dit voor de hand. We zijn nl. geïnteresseerd in het effect van de drie taakgerichte programma's en het controleprogramma zonder te generaliseren naar andere vergelijkbare programma's. Voor Factor A (scholen) ligt dit wat minder voor de hand. Het lijkt gewenst te kunnen generaliseren naar andere vergelijkbare scholen. De experimentele situatie op de scholen week echter af van de normale klasse-situatie. Zo kwamen er bijvoorbeeld instructeurs van buiten de school en werden de klassen verdeeld over vier groepen. Voorlopig lijkt het raadzaam geen generalisatie naar andere scholen na te streven maar eerst het effect te onderzoeken op de drie experimentele scholen.

Het design staat de toetsing toe van drie multivariate effecten nl. twee hoofdeffecten en een interactie-effect. Gezien de hypothesen is het hoofdeffect "programma's" (Factor B) het belangrijkste. In de eerste plaats om na te gaan of de programma's statistisch gezien effect hebben (de multivariate toets) en vervolgens om na te gaan of dit effect bereikt is op de test die correspondeert met het programma en of er sprake is van transfer-effecten van programma's op de scores van niet met die programma's corresponderende tests en op tests voor begrijpend lezen in het algemeen. De factor "scholen" dient er vooral toe om de aanwezigheid van interactie tussen scholen en programma's na te gaan. Bovendien wordt door het opnemen van deze factor voorkomen, dat verschillen tussen scholen in de error-term terecht komen. Interactie tussen scholen en programma's kan, gezien de aard van de verdeling van de instructeurs over de scholen en programma's, onder meer ontstaan door specifieke instructeurs-effecten.

De multivariate covariantie-analyses worden uitgevoerd met behulp van het programma MULTIVARIANCE (versie 6.2; Finn, 1977).

Bij de keuze van het aantal scholen werd mede rekening gehouden met het onderscheidingsvermogen van de uit te voeren statistische toetsen. Op basis van het feitelijke aantal leerlingen in de experimentele scholen kan nu het onderscheidingsvermogen worden uitgerekend. We beperken ons hier tot univariate F-toetsen voor het hoofdeffect "programma's". Bij een significantieniveau van 10% is het onderscheidingsvermogen in het geval van een middelgroot effect (vgl. Cohen, 1969) gelijk aan 0.89, hetgeen bevredigend is te noemen.

Aan multivariate covariantie-analyse ligt onder meer de vooronderstelling ten grondslag dat de regressiegewichten van de afhankelijke variabelen op de covariabele gelijk zijn voor alle cellen van het design. Om dit te toetsen werd een likelihood-ratio berekend die getransformeerd werd in een F-waarde (vgl. Finn, 1972). De multivariate toets voor paralleliteit van regressievlakken in de nameting heeft als F-waarde 0.96 ($p < 0.56$). Met de gegevens van de retentiemeting heeft deze grootte een waarde van 1.01 ($p < 0.45$). De hypothese van parallelle regressievlakken behoeft voor geen van beide analyses verworpen te worden, zodat covariantie-analyse gerechtvaardigd is.

Toetsing van de hypothesen: nameting

De gemiddelden en standaarddeviaties per cel staan voor de vijf tests vermeld in tabel c van Bijlage 5.5. Uit de gemiddelde waarden op de covariabele blijkt dat de verdeling van leerlingen over de vier groepen per school met het doel zo gelijk mogelijke groepen samen te stellen redelijk succesvol is geweest. De gemiddelde niveaus per school verschillen echter wel. De spreiding van de covariabele is in school 3 groter dan in de beide andere scholen. Ook op de andere variabelen kan deze tendens worden waargenomen. Er is evenwel geen sprake van duidelijke uitschieters wat betreft de standaarddeviaties per cel.

De resultaten van de multivariate covariantie-analyse van de gegevens van de nameting worden vermeld in tabel a van Bijlage 5.5. De multivariate F-toets voor het interactie-effect blijkt significant te zijn op het gekozen niveau van 10%. Aangezien echter geen enkele van de univariate F-toetsen significant is, wordt het interactie-effect genegeerd en wordt een model met twee hoofdeffecten geschat. Figuur 5.13. brengt de geschatte celgemiddelden

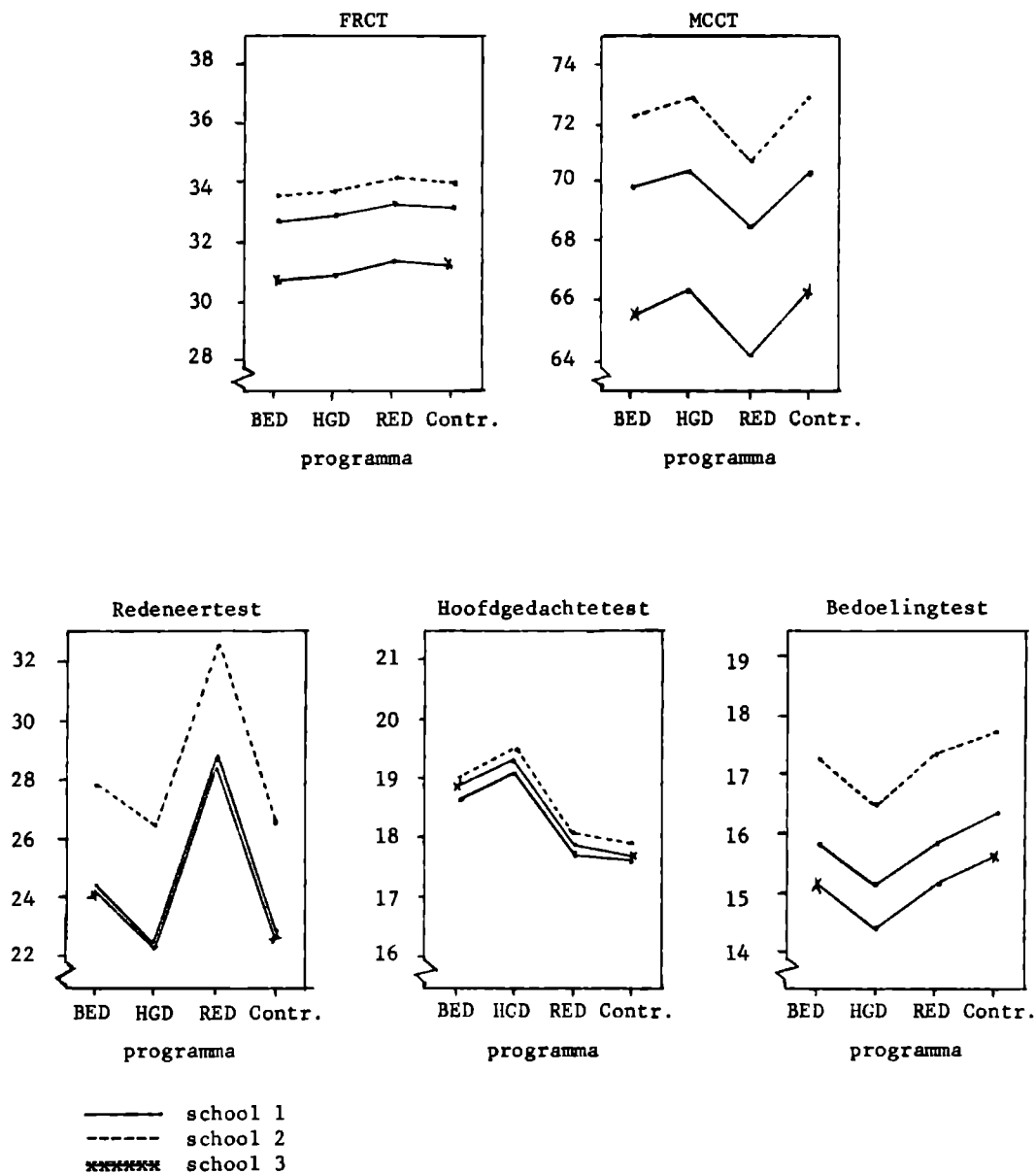
op basis van het hoofdeffecten covariantie-model in beeld (gegevens uit tabel d in Bijlage 5.5.). Door tabel c met tabel d in Bijlage 5.5. of met figuur 5.13. te vergelijken, kan worden nagegaan in hoeverre de geschatte waarden van de geobserveerde gemiddelden afwijken. De afwijkingen zijn over het algemeen gering. De grootste afwijking bedraagt nergens meer dan een halve standaarddeviatie. Dit bevestigt dat we met een interactie-effect geen rekening behoeven te houden. Een specifiek instructeurseffect behoeft op grond van dit resultaat niet te worden aangenomen.

De beide hoofdeffecten blijken bij multivariate toetsing significant te zijn. Hoewel sprake is van ongelijke celtaantallen (zie tabel 5.40.) en het design dus niet orthogonaal genoemd mag worden, is de correlatie tussen de contrasten van beide factoren zo gering ($r < 0.01$) dat bij de interpretatie van de hoofdeffecten geen rekening behoeft te worden gehouden met de volgorde van de beide factoren in dit design. Beide kunnen onafhankelijk van elkaar worden geïnterpreteerd. Het significante schooleffect lijkt grotendeels verklaard te kunnen worden door het feit dat de leerlingen van één school bij de voormeting lagere scores behaalden dan die van andere scholen.

Om de aard van het effect van de factor "programma's" te bepalen, kan vervolgens worden gekeken naar de univariate F-toetsen en naar de contrasteffecten binnen deze factor. In tabel 5.42. staan de geschatte effecten bij een vergelijking van de taakgerichte programma's met het controleprogramma. Met behulp van deze gegevens kunnen alle drie de hypothesen worden getoetst. Uit tabel 5.42. blijkt dat bij éézijdige toetsing ($p < 0.05$) de programma's Redeneren en Hoofdgedachte op de met die programma's corresponderende tests een significant effect hebben in vergelijking met het controleprogramma. Voor het programma Bedoeling geldt dit niet. Dit programma heeft geen positief effect op de Bedoelingtest, noch op enige andere test. Voor het programma Redeneren blijkt geen verschil met het controleprogramma op de tests die niet specifiek op de redeneertaak zijn afgestemd (tweezijdige toetsing). Het programma Hoofdgedachte verschilt significant van het controleprogramma op de Bedoelingtest. De controlegroep scoort gemiddeld hoger (ongeveer 1.2 punt) dan de hoofdgedachtegroep.

Het effect van elk taakgericht programma kan ook worden getoetst mede in vergelijking met de andere taakgerichte programma's. Bij de univariate F-toetsen binnen het hoofdeffect "programma's" wordt op één test een significant verschil tussen de instructiegroepen geconstateerd nl. op de Redeneer-

Figuur 5.13. Gemiddelde geschatte testscore per school in de nameting.
(Bijlage 5.5., tabel d).



290 Tabel 5.42. Schattingen van de, voor de factorscore uit de voormeting gecorrigeerde, contrasteffecten met de corresponderende standaardfouten op basis van het hoofdeffecten model. Nameting.

contrast	variabelen					standaardfouten				
	effect									
	FRCT	MCCT	HGD	RED	BED	FRCT	MCCT	HGD	RED	BED
Bedoeling- vs Controleprogramma	-0.436	-0.658	1.064	1.325	-0.506	0.940	2.052	0.860	1.019	0.685
Hoofdgedachte- vs Controleprogramma	-0.291	0.011	1.530 ^{xx}	-0.360	-1.224 ^x	0.929	2.028	0.850	1.007	0.677
Redeneer- vs Controleprogramma	0.121	-2.125	0.166	5.860 ^{xx}	-0.462	0.929	2.028	0.850	1.007	0.677

^x p < 0.10 (tweezijdige toetsing)

^{xx} p < 0.05 (éénzijdige toetsing)

test (zie tabel a in Bijlage 5.5.).

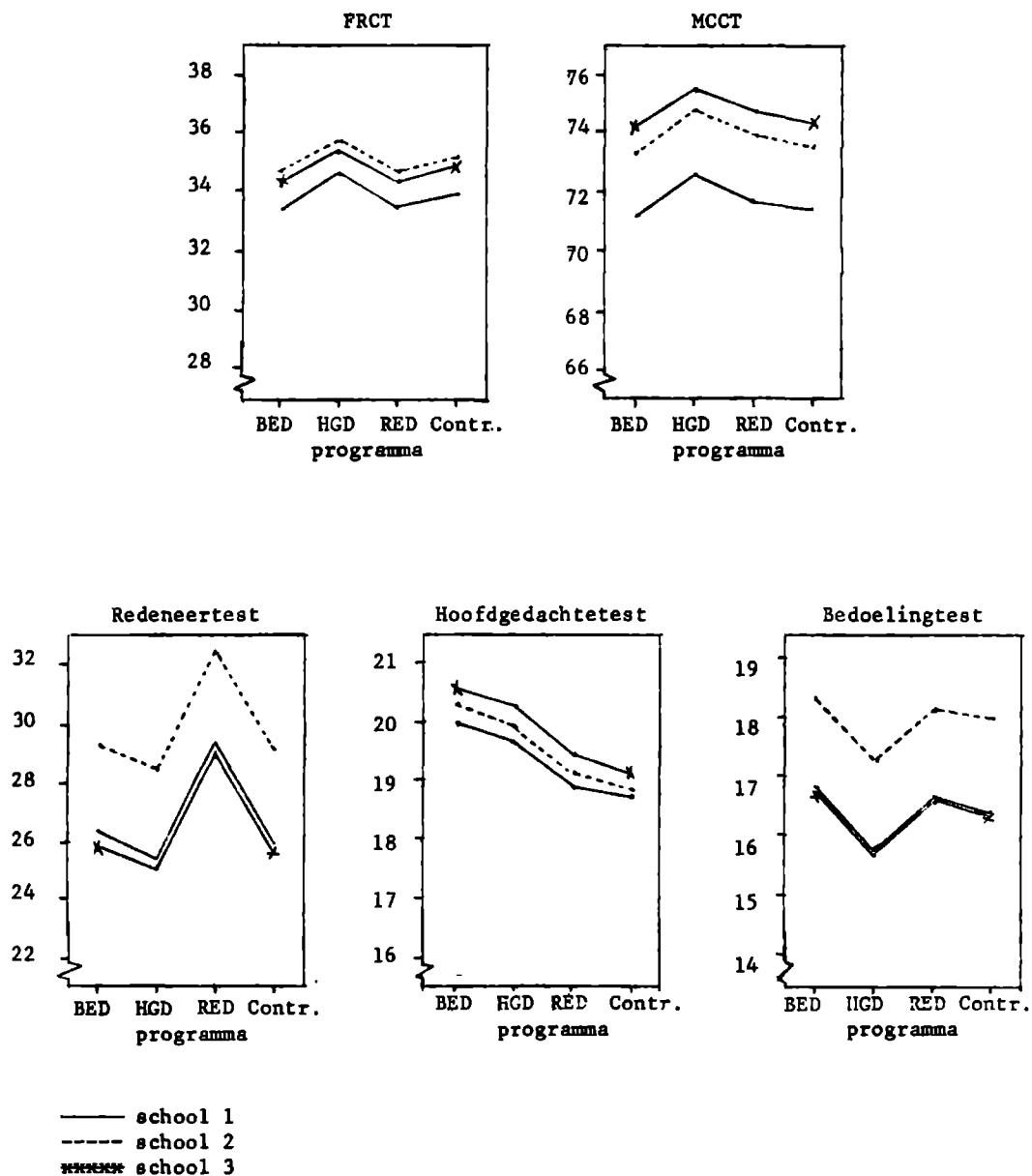
Uit de resultaten van deze eerste en tevens belangrijkste analyse blijkt dat, wat betreft het programma Redeneren, de hypothesen worden bevestigd. De hypothese dat het programma Redeneren een positief effect heeft in vergelijking met het controleprogramma wordt bevestigd (hypothese 1). De hypothese dat het programma geen transfer-effect heeft op tests die andere taken van begrip-pend lezen meten, wordt eveneens bevestigd (hypothese 2). Hieruit kan worden afgeleid dat de redeneertaak een empirisch onderscheidbaar aspect van begrip-pend lezen vormt. De grootte van het effect op de Redeneertest bedraagt, genomen ten opzichte van het gemiddelde van de drie andere groepen, iets meer dan één standaardafwijking (1.15 s). Tenslotte blijkt dat dit programma geen trans-fer-effect heeft op tests voor begrip-pend lezen in het algemeen (hypothese 3). De hypothesen met betrekking tot het programma Hoofdgedachte worden in deze analyse grotendeels bevestigd. Het programma Hoofdgedachte laat een positief effect zien op de Hoofdgedachtetest (ongeveer 0.3 s). De vraag naar de trans-fer-effecten laat zich niet éénduidig beantwoorden, vanwege het niet verwacht-te (overigens kleine) negatieve effect op de Bedoelingtest. Van positieve transfer is in ieder geval geen sprake, zodat ook de hoofdgedachtetaak als een empirisch onderscheidbaar aspect van begrip-pend lezen kan worden be-schouwd. De Hoofdgedachtetest brengt geen onderscheid aan tussen het program-ma Hoofdgedachte en Bedoeling, maar wel tussen deze beide samen en de andere twee programma's. Het gaat hier evenwel om een gering effect. De hypothesen betreffende het programma Bedoeling worden niet bevestigd. Omdat dit program-ma geen positief effect heeft op de daarmee corresponderende test, kan er geen uitspraak worden gedaan omtrent de onderscheidbaarheid van de bedoeling-taak en omtrent een mogelijk effect van het betreffende programma op tests die begrip-pend lezen in het algemeen meten.

Toetsing van de hypothesen: retentiemeting

Voor de analyse van de gegevens van de retentiemeting is hetzelfde design met dezelfde variabelen, condities en leerlingen gebruikt als bij de name-ting. Inplaats van de B-vorm is nu de A-vorm van de tests afgenomen.

De gemiddelden en standaarddeviaties per cel staan voor de vijf tests vermeld in tabel g van Bijlage 5.5. Vergelijken we deze tabel met de over-eenkomstige tabel voor de nameting (tabel c) dan blijkt dat de gemiddelde scores binnen elke cel in de retentiemeting over het algemeen hoger zijn dan in de nameting (twee maanden eerder).

Figuur 5.14. Gemiddelde geschatte testscore per school in de retentiemeting.
(Bijlage 5.5., tabel h).



Tabel 5.43. Schattingen van de, voor de factorscore uit de voormeting gecorrigeerde, contrasteffecten met de corresponderende standaardfouten op basis van het hoofdeffecten model. Retentiemeting.

contrast	effect					standaardfouten				
	variabelen									
	FRCT	MCCT	HGD	RED	BED	FRCT	MCCT	HGD	RED	BED
Bedoeling- vs Controleprogramma	-0.632	-0.201	1.365 [*]	0.137	0.372	0.783	2.246	0.811	1.179	0.700
Hoofdgedachte- vs Controleprogramma	0.612	1.103	1.071	-0.801	-0.686	0.774	2.220	0.801	1.166	0.692
Redeneer- vs Controleprogramma	-0.508	0.354	0.297	3.295 ^{**}	0.212	0.774	2.220	0.801	1.165	0.692

^{*} p < 0.10 (tweezijdige toetsing)

^{**} p < 0.05 (éénzijdige toetsing)

De resultaten van de multivariate covariantie-analyse van de gegevens van de retentiemeting worden vermeld in tabel f van Bijlage 5.5. De multivariate F-toets voor het interactie-effect blijkt niet significant te zijn. Figuur 5.14. brengt de geschatte celgemiddelden op basis van het hoofdeffecten covariantie-model in beeld. Door tabel g met tabel h in Bijlage 5.5. of met figuur 5.14. te vergelijken, kan worden nagegaan in hoeverre de geschatte waarden van de geobserveerde gemiddelden afwijken. Ook nu zijn de afwijkingen over het algemeen gering.

De beide hoofdeffecten zijn ook in de retentiemeting significant; deze effecten kunnen onafhankelijk van elkaar worden geïnterpreteerd.

In tabel 5.43. staan de geschatte effecten bij een vergelijking van de taakgerichte programma's met het controleprogramma. Uit deze tabel blijkt dat bij éénzijdige toetsing ($p < 0.05$) het programma Redeneren op de met dit programma corresponderende test in de retentiemeting ook een significant effect heeft in vergelijking met het controleprogramma. Het significant effect van het programma Hoofdgedachte is echter verdwenen. Bij éénzijdige toetsing heeft dit programma nog wel een significant effect op 10% niveau. Bij het programma Bedoeling wordt ook nu geen positief effect geconstateerd. Voor het programma Redeneren blijkt opnieuw geen verschil met het controleprogramma op de tests die niet specifiek op de redeneertaak zijn afgestemd (tweezijdige toetsing). Het programma Bedoeling verschilt significant van het controleprogramma op de Hoofdgedachtetest. De bedoelinggroep scoort hoger op de Hoofdgedachtetest dan de controlegroep. Er is blijkbaar een lichte tendens aanwezig dat de programma's Bedoeling en Hoofdgedachte van de beide andere programma's zijn te onderscheiden op de Hoofdgedachtetest.

Evenals bij de nameting kan ook nu het effect van elk taakgericht programma worden getoetst in vergelijking met de andere taakgerichte programma's. Bij de univariate F-toetsen binnen het hoofdeffect "programma's" wordt op één test een significant verschil tussen de instructiegroepen geconstateerd nl. op de Redeneertest (zie tabel f in Bijlage 5.5.).

Uit de resultaten van deze tweede analyse blijkt dat voor het programma Redeneren de drie eerder geformuleerde hypothesen worden bevestigd. De hypothesen met betrekking tot het programma Hoofdgedachte en Bedoeling worden niet bevestigd. Wat betreft het programma Hoofdgedachte is een tendens aanwezig in de richting van een klein positief effect in vergelijking met het controleprogramma.

Analyseprocedure

Het is mogelijk dat het effect van het controleprogramma in de experimentele scholen zodanig is geweest, dat daardoor de effecten van de experimentele programma's ten opzichte van het gangbare leesonderwijs onopgemerkt blijven. Om de realiteit van deze mogelijkheid te onderzoeken, zijn twee analyses uitgevoerd waarin naast leerlingen uit de experimentele scholen ook leerlingen uit de controlescholen zijn betrokken (vgl. par. 4.3.). De resultaten van deze analyses worden in deze paragraaf gepresenteerd.

Zoals reeds eerder vermeld, zijn de vijf ontwikkelde tests niet alleen op drie experimentele scholen afgenomen, maar ook op vier controlescholen (bij de retentiemeting op drie controlescholen). Op de controlescholen zijn geen speciale programma's door student-assistenten aangeboden. De leerlingen van deze scholen volgden het gangbare leesonderwijs zoals door de verschillende leerkrachten gerealiseerd. In deze paragraaf worden ongeveer dezelfde hypothesen getoetst als in de vorige paragraaf. Het verschil is dat nu geen vergelijking plaatsvindt tussen de prestaties van de leerlingen die een taakgericht programma hebben gevolgd en een controleprogramma, maar een vergelijking tussen de prestaties van de leerlingen van de drie experimentele programma's en het controleprogramma én de prestaties van de leerlingen die het gangbare leesonderwijs hebben gevolgd. Het volgen van het gangbare leesonderwijs van de klasse-leerkracht wordt als een aparte conditie in de analyse opgenomen. Het design zoals weergegeven in figuur 5.12. wordt hierdoor uitgebreid met vier scholen voor de nameting en drie scholen voor de retentiemeting.

Evenals in de vorige analyses worden de hypothesen getoetst met behulp van t-toetsen. Gekozen wordt voor een significantieniveau van 5% bij eenzijdige toetsing en van 10% bij tweezijdige toetsing. Deze toetsingen worden uitgevoerd in het kader van één enkelvoudige multivariate covariantie-analyse. Als covariabele wordt de factorscore gebruikt, die ook voor de leerlingen van de controlescholen is berekend op basis van tests die in de voormeting zijn afgenomen.

In de multivariate covariantie-analyse kunnen vijf condities worden onderscheiden nl. vier experimentele condities en één controleconditie i.c. het volgen van het gangbare leesonderwijs van de klasse-leerkracht. De laatste conditie wordt in de tabellen van deze paragraaf aangeduid als "geen ex-

tra programma" of "controlescholen". De celaantallen zijn in deze analyse ongelijk. De conditie "geen extra programma" bevat voor de nameting bijna zes maal en voor de retentiemeting bijna vijf maal zoveel leerlingen als de vier andere condities (zie tabel 5.40.).

Alvorens de covariantie-analyse tabel te interpreteren, dient nagegaan te worden of de hypothese van parallelle regressievlakken (of univariaat regressielijnen) niet wordt geschonden. De multivariate toets voor paralleliteit van regressievlakken heeft in de nameting een F-waarde van 1.17 ($p < 0.28$). Met de gegevens van de retentiemeting heeft deze grootte een waarde van 0.63 ($p < 0.89$). De hypothese van parallelle regressievlakken behoeft voor geen van de beide analyses verworpen te worden, zodat covariantie-analyse gerechtvaardigd is.

Omdat in deze analyse de vijfde groep leerlingen - de leerlingen van de controlescholen - niet op dezelfde wijze is samengesteld als de overige vier groepen, is het zinvol na te gaan in hoeverre deze groep vergelijkbaar is met de leerlingen van de experimentele scholen. In tabel j van Bijlage 5.5. is daartoe een overzicht opgenomen van gemiddelden en standaarddeviaties op alle tests uit de voormeting. De verschillen zijn zowel voor de gemiddelden als voor de standaarddeviaties redelijk klein en statistisch niet significant. Door toepassing van covariantie-analyse worden de eventuele verschillen in het niveau van begrijpend lezen uitgepartialiseerd.

Toetsing van de hypothesen: nameting en retentiemeting

De resultaten van de multivariate covariantie-analyse van de gegevens van de nameting worden vermeld in tabel k van Bijlage 5.5. De multivariate F-toets is zeer significant. Van de univariate toetsingsgrootheden is de F-toets voor de Free Response Cloze Test en voor de Redeneertest significant.

In tabel 5.44. staan de geschatte celgemiddelden van de afhankelijke variabelen na correctie van de covariabele. Uit deze tabel blijkt onder andere dat de leerlingen van de controlescholen (geen extra programma) gemiddeld hoger scoren op de Free Response Cloze Test dan de leerlingen van de experimentele scholen. Hetzelfde geldt, hoewel in mindere mate, voor de scores op de Multiple Choice Cloze Test.

Tabel 5.44. Geschatte celgemiddelden van de afhankelijke variabelen na correctie van de covariabele.

Nameting

conditie/test	FRCT	MCCT	HGD	RED	BED
programma Bedoeling	32.4	69.5	18.7	25.6	16.1
programma Hoofdgedachte	32.5	69.9	19.1	23.8	15.3
programma Redeneren	32.9	67.8	17.7	30.0	16.1
Controleprogramma	32.7	69.9	17.6	24.1	16.6
geen extra programma	34.2	70.6	18.5	25.5	15.6

In tabel 5.45. staan de geschatte effecten bij een vergelijking van de drie taakgerichte programma's en het controleprogramma met de controlescholen. Met behulp van deze gegevens kunnen de hypothesen worden getoetst. Uit tabel 5.45. blijkt dat bij éénzijdige toetsing ($p < 0.05$) het programma Redeneren een significant effect heeft op de met dit programma corresponderende test in vergelijking met de controlescholen. Voor de programma's Hoofdgedachte en Bedoeling geldt dit niet, al is er wel sprake van een tendens in de richting van een positief effect. Wat betreft het programma Redeneren is er geen verschil met de controlescholen op de Hoofdgedachtetest en de Bedoelingtest (tweezijdige toetsing). Hieruit kan worden afgeleid dat dit programma geen transfer-effect heeft op de Hoofdgedachte- en Bedoelingtest.

Zoals al eerder geconstateerd, scoren de leerlingen van de controlescholen significant hoger op de Free Response Cloze Test dan de leerlingen die de taakgerichte programma's en het controleprogramma hebben ontvangen. De leerlingen van de controlescholen scoren ook significant hoger op de Multiple Choice Cloze Test in vergelijking met de groep die het programma Redeneren heeft gevolgd. Hetzelfde geldt voor de Redeneertest bij de groepen die het hoofdgedachte- en controleprogramma hebben gevolgd. Opmerkelijk is dat de leerlingen die het controleprogramma hebben ontvangen significant hoger scoren op de Bedoelingtest dan de leerlingen van de controlescholen, terwijl op alle andere tests de controlescholen hoger scoren. Dit laatste doet vermoeden dat er overeenkomst bestaat tussen de inhoud van het controlepro-

Tabel 5.45. Schattingen van de, voor de factorscore uit de voormeting gecorrigeerde, contrasteffecten met de corresponderende standaardfouten. Nameting.

contrast	effect					standaardfouten				
	FRCT	MCCT	HGD	RED	BED	FRCT	MCCT	HGD	RED	BED
Bedoeling vs controlescholen	-1.852 [*]	-1.074	0.162	0.099	0.512	0.714	1.516	0.716	0.818	0.565
Hoofdgedachte vs controlescholen	-1.779 [*]	-0.639	0.611	-1.732 [*]	-0.304	0.702	1.489	0.703	0.803	0.555
Redeneren vs controlescholen	-1.389 [*]	-2.780 [*]	-0.747	4.482 ^{**}	0.469	0.701	1.488	0.702	0.803	0.555
Controle vs controlescholen	-1.511 [*]	-0.655	-0.914	-1.379 [*]	0.932 [*]	0.701	1.488	0.702	0.803	0.555

^{*} p < 0.10 (tweezijdige toetsing)

^{**} p < 0.05 (éénzijdige toetsing)

gramma en datgene wat de Bedoelingtest meet.

Uit de resultaten van deze analyse blijkt dat voor het programma Redeneren de drie hypothesen worden bevestigd. De hypothesen met betrekking tot het programma Hoofdgedachte en Bedoeling worden niet bevestigd. Wat betreft de laatstgenoemde programma's is er sprake van een tendens in de richting van een positief effect.

De resultaten van de multivariate covariantie-analyse van de gegevens van de retentiemeting worden vermeld in tabel 1 van Bijlage 5.5. De multivariate F-toets is significant. Van de univariate toetsingsgrootheden is de F-toets voor de Redeneertest significant.

In tabel 5.46. staan de geschatte celgemiddelden van de afhankelijke variabelen na correctie van de covariabele. Uit deze tabel blijkt onder andere dat de leerlingen van de controlescholen (geen extra programma) in bijna alle gevallen gemiddeld hoger scoren op de Free Response Cloze Test dan de leerlingen van de experimentele scholen. De verschillen zijn echter niet meer zo groot als in de nameting.

Tabel 5.46. Geschatte celgemiddelden van de afhankelijke variabelen na correctie van de covariabele.

Retentiemeting

conditie/test	FRCT	MCCT	HGD	RED	BED
programma Bedoeling	34.0	72.7	20.2	27.4	17.4
programma Hoofdgedachte	35.2	73.9	19.9	26.3	16.3
programma Redeneren	34.1	73.2	19.2	30.4	17.2
Controleprogramma	34.6	72.9	18.9	27.1	17.0
geen extra programma	35.1	72.5	19.3	26.4	16.9

In tabel 5.47. staan de geschatte effecten bij een vergelijking van de drie taakgerichte programma's en het controleprogramma met de controlescholen. Uit deze tabel blijkt opnieuw dat bij éénzijdige toetsing ($p < 0.05$) het programma Redeneren een significant effect heeft op de met dit programma corres-

Tabel 5.47. Schattingen van de, voor de factorscore uit de voormeting, gecorrigeerde contrasteffecten met de corresponderende standaardfouten. Retentiemeting.

contrast	variabelen					standaardfouten				
	effect									
	FRCT	MCCT	HGD	RED	BED	FRCT	MCCT	HGD	RED	BED
Bedoeling vs controlescholen	-1.193 [*]	0.172	0.952	0.977	0.578	0.665	1.785	0.695	0.928	0.579
Hoofdgedachte vs controlescholen	0.050	1.435	0.656	-0.089	-0.550	0.653	1.752	0.682	0.911	0.568
Redeneren vs controlescholen	-1.085 [*]	0.726	-0.109	4.004 ^{**}	0.353	0.653	1.752	0.682	0.911	0.568
Controle vs controlescholen	-0.577	0.373	-0.405	0.709	0.141	0.653	1.752	0.682	0.911	0.568

^{*} $p < 0.10$ (tweezijdige toetsing)

^{**} $p < 0.05$ (éénzijdige toetsing)

ponderende test in vergelijking met de controlescholen. Voor de programma's Hoofdgedachte en Bedoeling wordt, evenals in de nameting, geen significant effect gevonden. Het effect van deze programma's is overigens niet negatief. Wat betreft het programma Redeneren is er - met uitzondering van de Free Response Cloze Test - geen verschil met de controlescholen op de tests die niet specifiek op de redeneertaak zijn afgestemd (tweezijdige toetsing). Uit tabel 5.47. blijkt verder dat de leerlingen van de controlescholen significant hoger scoren op de Free Response Cloze Test in vergelijking met de groep die het programma Bedoeling heeft gevolgd.

Uit de resultaten van deze laatste analyse blijkt opnieuw dat voor het programma Redeneren de drie hypothesen worden bevestigd. De hypothesen met betrekking tot het programma Hoofdgedachte en Bedoeling worden niet bevestigd.

5.2.5. Samenvatting

In het experimentele gedeelte van het onderzoek worden drie hypothesen getoetst. De eerste hypothese heeft betrekking op het effect van drie taakgerichte programma's (Hoofdgedachte, Bedoeling, Redeneren) en is geformuleerd met het doel na te gaan in hoeverre taakgericht onderwijs op het gebied van het begrijpend lezen van teksten effectief is. Verwacht wordt dat de taakgerichte programma's een significant positief effect hebben op de met die programma's corresponderende tests (Hoofdgedachtetest, Bedoelingtest, Redeneertest) in vergelijking met een controleprogramma, dat niet op een specifieke leestaak is gericht. De tweede hypothese heeft betrekking op het effect van elk taakgericht programma op de beheersing van die taken waarop de beide andere programma's zijn gericht. Deze hypothese is opgesteld met de bedoeling een antwoord te geven op de vraag of de drie geselecteerde taken op het gebied van begrijpend lezen empirisch onderscheidbaar zijn. Verwacht wordt dat een taakgericht programma geen positief effect heeft op de beheersing van die taken waarop de beide andere programma's zijn gericht. De derde hypothese heeft betrekking op het effect van elk programma op de ontwikkeling van begrijpend lezen in het algemeen. Deze hypothese is geformuleerd met het doel na te gaan of taakgerichte programma's mogelijk een transfer-effect hebben op het begrijpend lezen in het algemeen, zoals gemeten met twee cloze tests. Verwacht wordt dat de taakgerichte programma's geen positief effect op de cloze tests hebben.

De drie hypothesen werden eerst getoetst in een analyse waarbij alleen de gegevens van de experimentele scholen werden gebruikt. Daarna werden ongeveer dezelfde hypothesen nogmaals getoetst in een analyse waarbij zowel van de gegevens van de experimentele scholen als van de controlescholen gebruik werd gemaakt.

De gegevens van de nameting en retentiemeting van de experimentele scholen werden geanalyseerd met behulp van multivariate covariantie-analyse. Hiervoor werd een twee factorieel design gebruikt met een indeling naar school en een indeling naar experimentele groep (drie taakgerichte programma's en een controleprogramma) als gekruiste factoren en de waarde van de factorscore uit de voormeting als covariabele.

De resultaten van de analyses in de nameting en retentiemeting vertonen in grote lijnen hetzelfde beeld. Hoewel in de nameting sprake is van een significant interactie-effect ($p < 0.10$) tussen scholen en groepen (programma's) behoeft hiermee geen rekening te worden gehouden. In de retentiemeting is geen sprake van een significant interactie-effect. In zowel de nameting als de retentiemeting zijn de twee multivariate hoofdeffecten - de indeling naar school en naar instructiegroep of programma - significant. Het significante schooleffect lijkt grotendeels verklaard te kunnen worden door het feit dat de leerlingen van één school op de voormeting lagere scores behaalden dan die van de andere scholen, terwijl bij latere metingen daar nog maar nauwelijks sprake van is.

Uit de contrasteffecten binnen de factor "programma's" blijkt dat de programma's Redeneren en Hoofdgedachte in de nameting een significant effect ($p < 0.05$) hebben op de met die programma's corresponderende tests. Deze programma's hebben geen positief transfer-effect op tests die andere taken van begrijpend lezen meten of op tests voor begrijpend lezen in het algemeen ($p < 0.10$). Uit de resultaten van deze analyse blijkt dat de drie hypothesen met betrekking tot de programma's Redeneren en Hoofdgedachte grotendeels worden bevestigd. Wat betreft het programma Bedoeling worden de hypothesen niet bevestigd.

In de retentiemeting heeft alleen het programma Redeneren een significant effect ($p < 0.05$) op de met dit programma corresponderende test. Het significant effect van het programma Hoofdgedachte is verdwenen. Bij éézijdige toetsing heeft dit programma nog wel een significant effect op 10% niveau. Voor het programma Redeneren bestaat er opnieuw geen verschil met het contro-

leprogramma op de tests die niet specifiek op de redeneertaak zijn afgestemd ($p < 0.10$). Uit de resultaten van deze analyse blijkt dat voor het programma Redeneren de drie hypothesen worden bevestigd. De hypothesen met betrekking tot het programma Hoofdgedachte en Bedoeling worden niet bevestigd. Wat betreft het programma Hoofdgedachte bestaat er wel een tendens in de richting van een positief effect.

De gegevens van de nameting en retentiemeting van de experimentele scholen en controlescholen samen werden geanalyseerd met behulp van een eenvoudige multivariate covariantie-analyse. Hiervoor werd de indeling naar experimentele groep (programma), die bij de experimentele scholen uit vier condities bestond, aangevuld met nog één conditie nl. het volgen van het gangbare leesonderwijs van de klasseleerkracht. (Deze onderzoeksopzet kan worden gekarakteriseerd als een quasi-experimenteel "non-equivalent control group design"). Als covariabele werd opnieuw de factorscore gebruikt.

De resultaten van de analyses van de gegevens van de nameting en retentiemeting vertonen in grote lijnen hetzelfde beeld. In zowel de nameting als de retentiemeting zijn de multivariate toetsen significant. Uit de vergelijking van de drie taakgerichte programma's en het controleprogramma met de controlescholen blijkt dat het programma Redeneren in de nameting en retentiemeting een significant effect ($p < 0.05$) heeft op de met dit programma corresponderende test. Wat betreft de programma's Hoofdgedachte en Bedoeling wordt in beide analyses geen significant effect gevonden, al is er bij de nameting en retentiemeting wel sprake van een tendens in de richting van een positief effect. Het programma Redeneren heeft opnieuw geen positief transfer-effect op tests die andere taken van begrijpend lezen meten of op tests voor begrijpend lezen in het algemeen ($p < 0.10$). Uit de resultaten van de analyses van de gegevens van de nameting en retentiemeting blijkt dat wat betreft het programma Redeneren de drie hypothesen worden bevestigd. De hypothesen met betrekking tot het programma Hoofdgedachte en Bedoeling worden bij deze onderzoeksopzet niet bevestigd.

Voor een totaaloverzicht van de geschatte directe effecten van de taakgerichte programma's verwijzen we naar tabel m van Bijlage 5.5.

Uit het bovenstaande blijkt dat de hypothesen met betrekking tot het programma Redeneren worden bevestigd. Het programma heeft zowel op korte als op lange termijn een positief effect, hetgeen betekent dat de redeneertaak

zoals die verwerkt is in het programma en zoals gemeten met de Redeneertest positief is te beïnvloeden. Verder heeft het programma geen positief transfer-effect op tests die andere taken van begrijpend lezen meten. Hieruit kan worden afgeleid dat de redeneertaak een empirisch onderscheidbaar aspect van begrijpend lezen vormt. Tenslotte blijkt - zoals voorspeld - dat het betreffende programma geen positief transfer-effect heeft op tests voor begrijpend lezen in het algemeen.

Uit het bovenstaande blijkt ook dat de hypothesen met betrekking tot het programma Bedoeling niet worden bevestigd. Omdat dit programma geen significant effect heeft op de daarmee corresponderende test, kan er geen uitspraak worden gedaan omtrent de onderscheidbaarheid van de bedoelings-taak en omtrent een mogelijk effect van het betreffende programma op tests die begrijpend lezen in het algemeen meten.

Uit het bovenstaande blijkt verder dat de hypothesen met betrekking tot het programma Hoofddedachte in één analyse grotendeels worden bevestigd en in andere analyses niet. Als wordt uitgegaan van de resultaten van die analyse die bij de sterkste onderzoeksopzet (het "true experimental pretest-posttest control group design") zijn verkregen, dan kan worden vastgesteld dat voor de korte termijn de hypothesen met betrekking tot het programma Hoofddedachte grotendeels worden bevestigd. Dit betekent in dat geval dat de hoofddedachtetaak op korte termijn positief is te beïnvloeden met behulp van het ontwikkelde programma en dat dit programma geen positief transfer-effect heeft op tests die andere taken van begrijpend lezen meten of op tests voor begrijpend lezen in het algemeen. Hieruit kan worden afgeleid dat de hoofddedachtetaak een empirisch onderscheidbaar aspect vormt van begrijpend lezen.

In het laatste hoofdstuk van deze studie worden eerst de belangrijkste resultaten van het onderzoek in het kort weergegeven en waar nodig van commentaar voorzien. Daarna worden enkele beperkingen van het onderzoek vermeld. Het hoofdstuk sluiten we af met aanbevelingen voor verder onderzoek.

6.1. Resultaten

Zoals reeds eerder vermeld, stonden in dit onderzoek drie vragen centraal nl. de vraag naar de betrouwbaarheid en validiteit van vijf ontwikkelde tests voor het begrijpend lezen van teksten, de vraag naar het effect van drie taakgerichte programma's op het gebied van begrijpend lezen en de vraag naar de realiteitswaarde van de onderscheiding van drie taken met betrekking tot begrijpend lezen.

Wat betreft de betrouwbaarheid en validiteit van de vijf paralleltests zijn de resultaten positief. Uit de hoogte van de coëfficiënten van stabiliteit, α -specificiteit en interne consistentie blijkt dat de betreffende tests voldoen aan de eisen die algemeen gesteld worden aan de betrouwbaarheid van meetinstrumenten. Met betrekking tot de validiteit van de vijf tests kan worden geconcludeerd dat de resultaten grotendeels overeenkomen met de verwachtingen. Een definitief antwoord op de vraag wat de verschillende tests meten, kan in het kader van dit onderzoek niet worden verwacht en gegeven. Verder onderzoek is hiervoor noodzakelijk. Dit laatste betekent overigens niet dat met de verspreiding van de betreffende tests moet worden gewacht. Gezien de kwaliteit van de tests lijkt het verantwoord deze op korte termijn ter beschikking te stellen van het onderwijs en het onderzoek. (De Hoofdgedachtetest wordt momenteel reeds gebruikt in het vervolgonderzoek, dat betrekking heeft op kenmerken van goede en zwakke lezers).

Wat betreft het effect van de drie taakgerichte programma's op het gebied van begrijpend lezen zijn de resultaten ten dele in overeenstemming met de verwachtingen. Als we ons beperken tot de resultaten van het experiment zoals uitgevoerd binnen de experimentele scholen, kan worden vastgesteld dat de programma's Redeneren en Hoofdgedachte bij de nameting een significant effect hebben. Hetzelfde geldt voor het programma Redeneren bij de retentieme-

ting, terwijl voor het programma Hoofddedachte een tendens in de richting van een significant effect wordt gevonden. Tegen de verwachting in heeft het programma Bedoeling geen significant effect bij de nameting en retentiemeting.

Op grond van deze resultaten is het niet mogelijk een algemene en definitieve uitspraak te doen omtrent de waarde van taakgericht onderwijs op het gebied van begrijpend lezen. Een dergelijke uitspraak zou niet in overeenstemming zijn met de resultaten en zou gezien het beperkte onderzoek naar deze vorm van leesonderwijs prematuur zijn. Een voorzichtige aanbeveling van taakgericht onderwijs op het gebied van begrijpend lezen lijkt evenwel geoorloofd. Het laatste betekent overigens niet dat het begrijpend lezen in de lagere school volledig zou moeten opgaan in taakgericht onderwijs. Naast de ontwikkeling, toepassing en integratie van kennis omtrent de wijze waarop een aantal belangrijke leestaken kunnen worden uitgevoerd, zal naar onze mening ook de nodige aandacht besteed moeten worden aan de ontwikkeling van meer algemene of gestructureerde kennis. Deze vorm van kennis is fundamenteel van aard en waarschijnlijk van groter belang voor het leren uitvoeren van bepaalde leestaken dan eerder werd vermoed. Het laatste zou een verklaring kunnen zijn voor het ontbreken van een significant effect van het programma Bedoeling bij de nameting en retentiemeting en van het programma Hoofddedachte bij de retentiemeting. Een nadere toelichting is hier noodzakelijk.

In de drie taakgerichte programma's is niet alleen aandacht besteed aan de ontwikkeling van gestructureerde kennis, maar overeenkomstig het uitgangspunt van de zogenaamde "skill-approach" ook en vooral aan de ontwikkeling van strategische kennis d.w.z. kennis met betrekking tot de wijze waarop de betreffende leestaken en hun onderscheiden deeltaken kunnen worden uitgevoerd. Uit de resultaten van het experiment, zoals uitgevoerd binnen de experimentele scholen, kan worden afgeleid dat de ontwikkeling van voornamelijk strategische kennis een positief effect heeft op de uitvoering van de redeneertaak en de hoofddedachtetaak (bij de nameting), maar niet effectief is bij de bedoelingtaak. Het zeer positieve effect op de redeneertaak (of Redeneertest) kan verklaard worden uit het feit dat bij deze taak vooral strategische of procedurele kennis van belang is. Gestructureerde kennis speelt bij deze taak een beperkte rol en kan zelfs negatief werken. Een verklaring

voor het ontbreken van een positief effect bij de bedoelingtaak (en gedeeltelijk ook bij de hoofdgedachtetaak) kan gezocht worden in het feit dat het programma er niet in geslaagd is de nodige strategische kennis bij de leerlingen te ontwikkelen. Hoewel deze verklaring niet kan worden uitgesloten, ligt een andere verklaring meer voor de hand nl. dat voor de uitvoering van de bedoelingtaak (en in mindere mate ook van de hoofdgedachtetaak) niet alleen strategische kennis van belang is maar ook en vooral gestructureerde kennis. Het vermoeden bestaat dat in het programma Bedoeling te weinig aandacht is besteed aan de ontwikkeling van gestructureerde kennis (kennis omtrent de conventionele tekststructuur en het conventionele verloop van doelgerichte handelingen), die nodig is om de bedoelingtaak (of Bedoelingtest) op een adequate wijze uit te voeren. Bij de hoofdgedachtetaak speelt gestructureerde kennis (kennis van concepten en dergelijke) vooral een belangrijke rol bij de Hoofdgedachtetest, waarbij een beroep wordt gedaan op kennis van begrippen. Het vermoeden bestaat dat een aantal leerlingen bij deze test met teksten is geconfronteerd die wat betreft de daarin voorkomende begrippen te moeilijk waren om op basis hiervan de hoofdgedachte af te leiden.

De bovengenoemde verklaring, die past binnen de schematheorie van onder anderen Rumelhart en Norman (1978), dient uiteraard op haar juistheid getoetst te worden. Indien mocht blijken dat ze geldig is, dan verandert ook de plaats en betekenis van taakgericht onderwijs, zoals hier opgevat. Een van de belangrijkste consequenties zou zijn dat deze vorm van leesonderwijs een beperkte functie heeft bij die leestaken, die naast strategische kennis ook en vooral een beroep doen op gestructureerde kennis. Aan de ontwikkeling van gestructureerde kennis dient in dat geval veel aandacht te worden besteed. Pas wanneer deze vorm van kennis voldoende is ontwikkeld, heeft het zin de nodige strategische kennis bij de leerlingen te ontwikkelen. De ontwikkeling van gestructureerde kennis kan worden gestimuleerd door de leerlingen veel goed geschreven teksten op hun (lees-)niveau te laten lezen en door tijdens de lessen voor begrijpend lezen (en ook bij andere lessen waar informatieverwerking via teksten centraal staat) veel aandacht te besteden aan de betekenis van concepten en de relatie tussen concepten.

Wat betreft de realiteitswaarde van de onderscheiding van de drie geselecteerde taken op het gebied van het begrijpend lezen van teksten blijkt uit het experiment, zoals uitgevoerd op de experimentele scholen, dat het

programma Redeneren in de nameting en retentiemeting geen positief transfer-effect heeft op tests die andere taken van begrijpend lezen meten. Hetzelfde geldt voor het programma Hoofdgedachte in de nameting. Uit deze gegevens kan worden afgeleid dat de redeneertaak en de hoofdgedachtetaak empirisch onderscheidbare aspecten van begrijpend lezen vormen. Omdat het programma Bedoe-ling geen significant effect heeft op de daarmee corresponderende test, kan er geen uitspraak worden gedaan omtrent de onderscheidbaarheid van de bedoe-lingtaak. Uit de resultaten van een analyse, die in het descriptieve gedeel-te van het onderzoek is uitgevoerd, blijkt in ieder geval dat er ruimte is voor de veronderstelling dat ook deze taak een specifiek aspect van begrij-pend lezen vormt.

De bovengenoemde resultaten ondersteunen de legitimiteit van taakgericht onderwijs (de zogenaamde "skill-approach") op het gebied van begrijpend le-zen en ontkrachten enkele uitgangspunten van de "holistic approach" van Goodman (1972), Smith (1978) en Vacca (1980). Uit de resultaten blijkt dat de unifactor theorie (vgl. Chapman, 1971), die postuleert dat begrijpend le-zen uit één enkelvoudige algemene vaardigheid bestaat, waaraan geen deel-vaardigheden of taken kunnen worden onderscheiden, verworpen moet worden. De resultaten bevestigen in grote lijnen wat Davis (1968, 1972) reeds in eerder onderzoek, dat niet experimenteel van aard was, heeft gevonden.

6.2. Beperkingen van het onderzoek

Bij de beoordeling van de resultaten van onder andere het experimentele gedeelte van het onderzoek dient rekening te worden gehouden met de beper-kingen van het onderzoek. Deze beperkingen houden verband met de onvolkomen-heden in de opzet van het onderzoek. In hoofdstuk IV hebben we de onder-zoeksoptzet binnen de experimentele scholen gekarakteriseerd als een "true experimental pretest - posttest control group design". De opzet waarbij zo-wel de experimentele als de controlescholen betrokken zijn, hebben we gety-peerd als een quasi-experimenteel "non-equivalent control group design".

De onvolkomenheden van deze designs verzwakken, wat Campbell en Stanley (1971) noemen, de interne en externe validiteit van het onderzoek. De inter-ne validiteit betreft de interpreteerbaarheid van de resultaten: had de ex-perimentele variabele werkelijk invloed op de uitkomsten of zijn deze uit-komsten te beschouwen als artefacten van de procedure? De externe validiteit

heeft betrekking op de generaliseerbaarheid van de uitkomsten: naar welke populaties, gelegenheden en metingen kunnen de gevonden resultaten geïnfereerd worden? Naast de interne en externe validiteit onderscheiden Cook en Campbell (1976) nog twee vormen van validiteit nl. statistische conclusievaliditeit en constructvaliditeit. In het volgende zullen we ons beperken tot de interne en externe validiteit en de vraag naar deze vormen van validiteit voor het betreffende onderzoek aan de orde stellen aan de hand van de eerdere publicatie van Campbell en Stanley (1971).

Campbell en Stanley (1971) onderscheiden acht factoren, die de interne validiteit van een onderzoeksopzet kunnen aantasten nl. historiciteit, rijping, testeffect, instrumentatie, statistische regressie, selectie, uitval en interactie-effecten van selectie met rijping, selectie met getest worden en dergelijke. De betreffende auteurs onderscheiden vier factoren, die de externe validiteit kunnen aantasten. De volgende drie factoren zijn in dit verband van belang nl. interactie-effect van getest worden en experimentele variabele, interactie-effecten tussen selectie en experimentele variabele, en reactieve effecten in de experimentele opzet.

Met betrekking tot de interne validiteit van het onderzoek opgezet als een "true experimental pretest - posttest control group design" kan worden opgemerkt dat factoren als historiciteit, rijping, testeffect, instrumentatie en dergelijke zeer waarschijnlijk niet als rivaliserende verklaringsmogelijkheden in aanmerking komen. Een nadere toelichting op enkele belangrijke factoren is hier op zijn plaats.

Onder historiciteit verstaan Campbell en Stanley (1971) bijzondere externe gebeurtenissen tussen voormeting, nameting en retentiemeting die eerder dan de experimentele variabele het gevonden effect kunnen verklaren. Onder deze term verstaan zij ook de invloeden die de experimentele groep en de controlegroep afzonderlijk ondergaan (intrasession history). Met betrekking tot historiciteit kan worden opgemerkt dat verschillende maatregelen zijn genomen om deze factoren zo veel mogelijk onder controle te houden. Om te voorkomen dat een mogelijk instructeurseffect versmolten zou raken met één programma of één school werd besloten dat de acht instructeurs op verschillende experimentele scholen verschillende programma's zouden geven. Indien sprake was geweest van een instructeurseffect zou dit in de analyse moeten blijken uit een significant interactie-effect tussen scholen en programma's.

Hoewel uit de analyse bleek dat in één meting sprake was van een significant interactie-effect, behoefde hiermee geen rekening te worden gehouden. Ter voorkoming van een mogelijk intrasession-effect werden de programma-instructeurs evenals de testinstructeurs met opzet onkundig gehouden van het feit dat één van de groepen als controlegroep fungeerde en één van de programma's als controleprogramma. Van een compenserende rivaliteit of ontmoediging bij de controlegroep kan dan ook nauwelijks sprake zijn (vgl. Cook & Campbell, 1976). Om ongewenst oefenen tijdens of na de experimentele procedure te voorkomen, werd aan de leerkrachten van de betreffende vierde klassen geen informatie verstrekt over de inhoud en de functie van de verschillende programma's en was het voor hen niet mogelijk de lessen bij te wonen. Met de leerkrachten werd afgesproken om tijdens en na de experimentele procedure op dezelfde wijze leesonderwijs te geven als voorheen. Het idee om tijdens de experimentele procedure het leesonderwijs voor de betreffende vierde klassen op een of andere manier onder controle te houden, is wel overwogen, maar om praktische redenen niet uitgevoerd. De mogelijkheid dat de betreffende leerkrachten tijdens en na de experimentele procedure extra aandacht hebben besteed aan de drie onderscheiden taken kan niet geheel worden uitgesloten gezien het feit dat zij via de testafnames op de hoogte waren van de inhoud van de verschillende tests. Deze factor kan echter nauwelijks als een rivaliserende verklaringsmogelijkheid voor het effect van het programma Redeneren en Hoofdgedachte in aanmerking komen, gezien het vrij specifieke karakter van de redeneertaak en de hoofdgedachtetaak.

Wat betreft de invloed van het testen en de instrumentatie zijn hier twee opmerkingen van belang. De eerste opmerking heeft betrekking op de testinstructeurs. Zoals uit hoofdstuk IV blijkt, zijn tijdens het experiment alle tests op een school door dezelfde testinstructeur afgenomen. De mogelijkheid dat hierdoor een testinstructeur-effect is opgetreden, welke versmolten is met de school, is hierbij niet geheel uitgesloten. Controle heeft hierop niet plaatsgevonden. De tweede opmerking betreft de volgorde en het tijdstip van de tests. In de voormeting zijn de tests op alle scholen in dezelfde volgorde afgenomen, terwijl in de nameting en retentiemeting de volgorde van afname systematisch is gevarieerd. Voor de experimentele scholen betekende dit dat bijvoorbeeld de Hoofdgedachtetest in de ene school als eerste en in de andere school als laatste test (ongeveer een maand na de laatste les van het pro-

gramma Hoofdgedachte) werd afgenomen. Bij de analyse van de gegevens is met deze verschillen tussen laatste les en afnamemoment geen rekening gehouden. De mogelijkheid is niet uitgesloten dat de gevolgde testprocedure van invloed is geweest op het effect van de drie taakgerichte programma's in de nameting en retentiemeting.

Wat betreft de selectie kan worden opgemerkt dat binnen elke experimentele school gestreefd is naar een zo gelijk mogelijke samenstelling van de experimentele groep en controlegroep. Na matching op leesbekwaamheid werden de proefpersonen op basis van toeval over de groepen verdeeld. Hoewel gesproken kan worden van randomtoewijzing van proefpersonen aan condities, is het van belang op te merken dat in dit onderzoek geen volledige randomisering is toegepast. De randomisering kende twee restricties; zij vond plaats binnen elke experimentele school en binnen de door de matchingsprocedure gestelde grenzen.

Met betrekking tot de externe validiteit van het onderzoek opgezet als een "true experimental pretest - posttest control group design" kan het volgende worden opgemerkt.

De vraag in hoeverre er sprake is van een interactie-effect tussen de voormeting en de experimentele variabele kan hier moeilijk worden beantwoord, omdat niet voor een dergelijk effect is gecontroleerd. De mogelijkheid dat de taakgerichte tests uit de voormeting de leerlingen van de experimentele groep extra gevoelig hebben gemaakt voor de taakgerichte programma's kan niet worden uitgesloten, al is de kans dat dit werkelijk een rol heeft gespeeld gering gezien de vele andere tests die in de voormeting zijn afgenomen.

Een interactie tussen selectie en experimentele procedure behoeft bij deze opzet nauwelijks te worden gevreesd. Bij het aanzoeken van de deelnemende scholen deden zich geen problemen voor; alle aangezochte scholen bleken bereid te zijn aan het onderzoek deel te nemen. Er zijn geen redenen om aan te nemen dat de deelnemende scholen op karakteristieke kenmerken van andere scholen verschillen; karakteristieke kenmerken die samen zouden kunnen hangen met de experimentele condities. Het is duidelijk dat de experimentele scholen, evenals de experimentele en controlescholen samen, geen aselechte steekproef vormen uit de landelijke populatie scholen voor gewoon lager onderwijs. Dit betekent dat de resultaten van het experimente-

le gedeelte, waarbij twee designs worden gebruikt, alleen generaliseerbaar zijn naar de verzameling "grote scholen voor gewoon lager onderwijs van het voormalig stadsgewest Nijmegen, inclusief de gemeente Elst".

Reactieve effecten in de experimentele opzet vormen eveneens een bedreiging van de externe validiteit. Een onnatuurlijke experimentele setting en veelvuldig testen kunnen bijvoorbeeld leiden tot veranderingen in de subjecten zelf en daardoor de uitkomsten van het experiment beïnvloeden. De vraag kan gesteld worden of met name de experimentele setting van het onderzoek en het veelvuldig afnemen van tests reactieve effecten hebben veroorzaakt. Het feit dat de leerlingen in de experimentele setting in vier groepen werden verdeeld en in verschillende lokalen onderwijs kregen van "vreemde" leerkrachten kan reactieve effecten hebben opgeroepen. De eerste twee lessen voorafgaande aan de programma's waren bedoeld om mogelijke reactieve effecten "op te lossen". Verder moet worden opgemerkt dat de leerlingen van de betreffende experimentele scholen gewend waren om onderwijs te krijgen van "vreemde" leerkrachten i.c. studenten van pedagogische academies. De kans dat de experimentele setting reactieve effecten heeft opgeroepen, is ons inziens niet groot. De kans dat het veelvuldig testen reactieve effecten heeft gehad, is daarentegen groter. Het feit dat een aantal gegevens van één experimentele school in de nameting niet betrouwbaar bleek te zijn, kan wijzen in de richting van reactieve effecten door het veelvuldig afnemen van tests.

Samenvattend kan worden geconcludeerd dat de resultaten van het onderzoek uitgevoerd op de experimentele scholen onderhevig zijn aan enkele beperkingen met name wat betreft de externe validiteit.

Met betrekking tot de interne en externe validiteit van het onderzoek waarbij zowel de experimentele als de controlescholen betrokken zijn, kunnen we kort zijn. Zoals reeds eerder vermeld, is de onderzoeksopzet te typeren als een quasi-experimenteel "non equivalent control group design".

Wat betreft de interne validiteit zijn twee factoren hier van belang nl. historiciteit en selectie. Wat betreft de historiciteit kan worden opgemerkt dat er geen maatregelen zijn genomen om deze factor voor de controlegroep (de controlescholen) onder controle te houden. De mogelijkheid dat de controlegroep in dit geval verschillende invloeden heeft ondergaan is zeker niet uitgesloten. Hoewel uit de vragenlijst niet bleek dat het leesonderwijs van

de controlescholen sterk verschilde, is de mogelijkheid van verschillen de invloeden niet uitgesloten.

Wat betreft de selectie kan het volgende worden opgemerkt. Op grond van de beperkte steekproef kon niet worden verwacht dat de experimentele groep en de controlegroep op relevante variabelen gelijk zouden zijn. Uit de scores op de tests die in de voormeting werden verzameld, bleek dat de verschillen tussen beide groepen redelijk klein waren. Door toepassing van covariantie-analyse zijn de eventuele verschillen in niveau van begrijpend lezen uitgepartialiseerd.

Wat betreft de externe validiteit kunnen bij deze opzet ongeveer dezelfde opmerkingen worden gemaakt als bij de voorgaande onderzoeksopzet.

Samenvattend kan worden geconcludeerd dat de resultaten van het onderzoek, waarbij zowel de experimentele als de controlescholen betrokken zijn, aan een aantal beperkingen onderhevig zijn die betrekking hebben op zowel de externe als de interne validiteit van het onderzoek.

6.3. Verder onderzoek

Het is een goed gebruik een onderzoek af te sluiten met een aantal aanbevelingen voor verder onderzoek.

Wat betreft de vijf ontwikkelde tests is verder onderzoek gewenst. In de eerste plaats dient nagegaan te worden of de betreffende tests ook voor de hogere leerjaren van de lagere school geschikt zijn. Uit het vervolgonderzoek, dat momenteel wordt uitgevoerd, blijkt bijvoorbeeld dat de Hoofdgedachtetest ook in het vijfde leerjaar bruikbaar is. In de tweede plaats dient wat betreft de betrouwbaarheid i.c. de homogeniteit van de drie taakgerichte tests te worden onderzocht in hoeverre deze tests voldoen aan de eisen van het twee parameter logistische latente trek model en aan de meer restrictieve eisen van het Rasch-model. Voor een toepassing van deze modellen op eerder door ons ontwikkelde hoofdgedachtetests verwijzen we naar Buitenhuis (1981). Wat betreft verder onderzoek naar de validiteit van de vijf ontwikkelde tests denken we aan onderzoek waarbij de samenhang wordt vastgesteld tussen de betreffende tests en taaltests, verbale intelligentietests, tests voor begrijpend lezen (C.I.T.O., 1980), leesattitudeschalen en conceptuele kennistests. Ook dient onderzocht te worden of de items van de

taakgerichte tests inderdaad tekstafhankelijk zijn (vgl. Tuinman, 1973-1974). Verder zal in een exploratief onderzoek moeten worden nagegaan welke oplossingsstrategieën leerlingen hanteren bij het uitvoeren van de drie taakgerichte tests en de twee cloze tests. Hierop aansluitend kan bij de taakgerichte tests worden vastgesteld wat het verschil in prestatie is tussen het selecteren van het juiste antwoord uit alternatieven en het genereren of zelf formuleren van het juiste antwoord (vgl. Van Blaricom & White, 1975).

Wat betreft het effect van de programma's en de realiteitswaarde van de drie onderscheiden taken is een replicatie-onderzoek van belang. Nagegaan kan worden wat het effect is van bijvoorbeeld het programma Hoofddedachte in leerjaar 4 en 5 en wat het transfer-effect is van dit programma op de uitvoering van de redeneertaak en de bedoelingtaak. Hierbij kan aan het programma Hoofddedachte meer tijd (lessen) worden besteed en kunnen ook tijdens de uitvoering van het programma enkele tests worden afgenomen. Te denken valt aan tests met betrekking tot het identificeren van het thema in een tekst en het identificeren van de hoofddedachte (topic sentence). Om na te gaan wat de invloed van gestructureerde kennis is bij de uitvoering van de hoofddedachtetaak (i.c. de Hoofddedachtetest) kan een conceptuele kennistest worden ontwikkeld, die de belangrijkste begrippen bevat, die in de teksten van de Hoofddedachtetest voorkomen. Deze conceptuele kennistest zal, indien de verklaring juist is dat gestructureerde kennis een belangrijke rol speelt bij de uitvoering van de hoofddedachtetaak, een hoge voorspellende waarde moeten hebben ten aanzien van de Hoofddedachtetest.

Een andere aanbeveling voor verder onderzoek is meer fundamenteel van aard en betreft de noodzaak in de toekomst meer onderzoek te doen naar het proces van begrijpend lezen bij leerlingen van de lagere school en naar het effect van interventies in de vorm van specifieke didactische werkwijzen en materialen. Bovendien zal meer onderzoek moeten worden gedaan naar de praktijk van het begrijpend leesonderwijs in de lagere school. Onderzoek naar het proces van begrijpend lezen en naar het effect van interventies kan belangrijke aanwijzingen geven voor de optimalisering van het begrijpend leesonderwijs. Het vervolgonderzoek, dat momenteel wordt uitgevoerd en betrekking heeft op zwak begrippele lezers uit leerjaar 3 en 5, is op beide aspecten gericht. Kennis van het proces van begrijpend lezen en van het effect van interventies is een noodzakelijke maar niet voldoende voorwaarde

voor de optimalisering van het begrijpend leesonderwijs. Kennis omtrent de praktijk van het begrijpend leesonderwijs is hiervoor eveneens noodzakelijk (vgl. Goodlad et al., 1970, 1979; Velema, 1966). Onderzoek naar de praktijk van het begrijpend leesonderwijs is, zoals Durkin (1977, 1978-1979) opmerkt, zeer schaars en voor degenen die de praktijk niet kennen ontvullend. Om een dergelijk onderzoek naar de praktijk van het begrijpend leesonderwijs uit te voeren, kan gebruik worden gemaakt van een herziene versie van de vragenlijst zoals in dit onderzoek ontwikkeld en van de observatie-categorieën die Durkin in haar onderzoek hanteert.

ASPECTS OF READING COMPREHENSION IN THE FOURTH GRADE OF THE ELEMENTARY SCHOOL

The research reported in this thesis fits into a broad research plan and is a follow up to previously carried out research. Our aim is to contribute to the knowledge of curriculum development in reading comprehension.

The Wisconsin Design for Reading Skill Development, an American reading curriculum based on the skill-approach, inspired this investigation. The most characteristic aspect of the skill-approach is that a series of reading skills is used as a starting point and that in reading instruction systematic and often separate attention is paid to the development of the various skills. One advantage of this approach is that reading comprehension instruction can proceed in a direct and systematic manner; the teacher can focus his attention on a particular skill and teach the pupils how such a skill can be acquired. Moreover, reading comprehension instruction can be differentiated. Another advantage is that by using specific tests the teacher can check whether and to what extent pupils have mastered a specific reading skill. Furthermore shortcomings in the different curriculum units or in the teaching learning situations can be traced. Besides these advantages the skill-approach also bears certain problems. Between the advocates of this approach there is a lack of agreement as to what is meant by the term "skills". Furthermore the advocates do not base their skills sufficiently on a theory of reading comprehension. In addition the postulated skills are often formulated in vague behavioral terms omitting the characteristics of the object of the behavior i.e. the text. Another problem are the tests which with respect to validity and reliability often can be improved. Up to now little is known about the effect of this approach on the mastery of different reading skills and on general reading achievement. A fundamental problem of the skill-approach is whether reading comprehension is a related set of skills or one general skill with indistinguishable subskills. This last problem is directly related to the legitimacy of the skill-approach to reading comprehension. Although still other shortcomings of this approach can be indicated, we have confined ourselves to those mentioned above. Here-

by we assume that the skill-approach can contribute positively to the planning and carrying out of teaching reading comprehension in the elementary school. Since the term "skills" refers to goals, activities and processes, we prefer to use the term "tasks" and "task directed instruction" in our research. Tasks are regarded as instructions which the pupils must learn to carry out with respect to a text. By task directed instruction in reading comprehension we mean the method of teaching whereby the teacher pays explicit attention to one reading task or a number of connected reading tasks and by which pupils learn step by step how such a task or a number of connected tasks can be carried out. The focus is here on the development of strategic or procedural knowledge.

The most important objectives of our research can be summarized as follows:

- (a) to determine the reliability and validity of five tests with respect to reading text comprehension;
- (b) to determine the effect of three different experimental programs (learning units), each of which is directed to an improved mastery over a special task regarding reading text comprehension;
- (c) to determine the validity of the distinction of three tasks on reading comprehension.

With respect to the tasks mentioned in (b) and (c), fourth grade pupils must infer information from a text. These tasks are called inferential reading tasks.

In the first stage of our research a literature study was carried out dealing with the three most important research questions. The results of this study were of importance not only for constructing the five tests and the three task directed programs, but also for the design of the study. According to Chapman (1971) we decided to investigate the validity of the distinction of the three inferential reading tasks via an experimental design.

Then we attempted to select three inferential reading tasks on the basis of a theory of reading comprehension and inferential processes. Our experience was that theories of reading comprehension have not yet sufficiently been worked out or tested in order to select inferential reading tasks on this basis. Nevertheless we inferred and formulated some reading tasks

from these theories. The following tasks were selected: inferring the main idea from an informative text, inferring the intention of the principal character from fictitious narrative and drawing inferences from logical reasonings. Once the tasks had been selected, we tried to specify and justify the tasks for which tests and programs had to be developed. With regard to the characteristics of the texts, the tasks could be fairly accurately described. However, it appeared not to be possible to describe the processes which play a role in executing the reading tasks, because a coherent theory on reading comprehension did not exist.

The tasks were operationalized in the form of three task directed tests, namely the Main Idea Test, the Intention Test, and the Reasoning Test. In constructing these tests every item was based on a separate text according to Davis (1968). For each test try-outs were made on different groups of pupils from the fourth grade. Since the tests had to be administered several times by the same pupils parallel tests were developed. The same applied for two types of cloze tests namely the Free Response Cloze Test and the Multiple Choice Cloze Test. For each of the three selected tasks we developed experimental programs (units) whereby the Gagné-procedure (1977b) was broadly followed. Each program was directed to master an inferential reading task. The programs consisted of about ten lessons, which contrary to the Wisconsin Design were developed for classroom usage. The Main Idea, Intention and Reasoning program were tested on a number of fourth grades and improved where necessary. With the task directed tests a first indication could be obtained of the effect of each program on mastering the reading task involved.

With regard to the design and methodology of the research it is important to note the following. Eight large elementary schools were selected systematically and divided at random into a group of four experimental schools and into another group of four control schools. The fourth grade of the first group of schools was presented with the three experimental programs and a control program, while the fourth grade of the second group of schools received no special program. During the pre-testing ($N = ca. 500$) five standardized tests on decoding, reading comprehension, vocabulary and non-verbal intelligence, the three task directed tests and both cloze tests were administered in all schools. During the post-testing and

retention-testing the two kinds of tests last mentioned or parallel forms were again administered in the same schools. It should be mentioned that the data collected in the pre-testing ($N = \text{ca. } 500$) and in the post- and retention-testing of the control schools ($N = \text{ca. } 220$) were of importance in connection with the reliability and validity of the tests developed.

On the basis of a factor score from the pre-testing the pupils ($N = 182$) from the fourth grades of each experimental school were matched on the basis of reading ability and divided at random into four groups. For a period of five weeks each group received instruction in one of the three experimental programs or a control program. The last mentioned program was related to reading comprehension but not to the three reading tasks selected. As stated each program consisted of 10 lessons and was given by graduate students. The design of the experiment in the four experimental schools can be characterized as a "true experimental pretest-posttest control group design" (Campbell & Stanley, 1971). Besides this design another design was possible, whereby the effect of the three experimental programs and the control program will be compared with the effect of the normal reading instruction as carried out in the control schools. This design can be regarded as being a quasi-experimental "non-equivalent control group design" (Campbell & Stanley, 1971).

The second stage of our research deals with the analysis of the data and the results of the research. This stage consists of two separate sections, namely a descriptive section and an experimental section.

The main objective of the descriptive section was to determine the reliability and validity of the five parallel tests developed. In order to define the reliability of the tests four methods (test-retest, parallel-form, splithalf and item-analysis) were applied. The results were satisfactory with regard to the reliability of the five tests. From the size of the coefficients of stability, a-specificity and internal consistency, it appeared that both the task directed tests and the cloze tests come up to the requirements of reliability. With regard to the construct validity of the five tests, it can be concluded that to a great extent the results coincide with our expectations. The intercorrelations between the five tests were substantial ($.60 < r < .70$). As expected the Multiple Choice Cloze Test appeared not to be a measure for a specific aspect of reading compre-

hension, namely literal comprehension. Furthermore the correlation between the task directed tests was low ($r < .40$) with a measure for decoding. With respect to the cloze tests a moderate ($.40 < r < .60$) correlation was found with this criterium. In general the five tests correlated to a substantial degree with a measure for reading comprehension. The relationship between the five tests and a measure for reading vocabulary varied from substantial ($.60 < r < .70$) to high ($r < .70$). As expected the five tests showed only a moderate relationship with a measure for non-verbal intelligence. The scores of the tests developed correlated higher with the grade point average for reading comprehension than with the grade point average for decoding. The correlations between these tests and the marks for reading comprehension were generally speaking higher than the correlations between these tests and the marks for language arts and arithmetic.

In the experimental section of our study three hypotheses were tested. The first hypothesis dealt with the effect of the three task directed programs. These programs were expected to have a significant effect on the tests corresponding to these programs in comparison with the control program. The second hypothesis dealt with the question as to whether the three tasks selected were empirically distinctive. A task directed program was expected to have no significant effect on the ability to master the tasks to which both other programs were directed. The third hypothesis dealt with the effect of each program on the development of reading comprehension in general. The task directed programs were expected to have no significant effect on the cloze tests developed.

First data from the experimental schools were used. Then almost the same hypotheses were tested again in an analysis whereby data from both experimental schools and control schools were used. The data were analysed by means of multivariate covariance-analysis.

The analyses showed that the hypotheses related to the Reasoning program were confirmed. The program had a significant effect in both post-testing and retention-testing ($p < 0.05$), which indicates that the reasoning task as incorporated in the program and as measured by the Reasoning Test can be positively influenced. As predicted the program had no positive transfer-effect on tests which measure other tasks of reading comprehension ($p < 0.10$). From this we can deduce that the reasoning task is an empirical

distinctive aspect of reading comprehension. Finally it appeared that this program had no positive transfer-effect on tests which measure reading comprehension in general.

Further, the analyses showed that the hypotheses related to the Main Idea program were confirmed in one analysis and not in other analyses. Confining ourselves to the results of the analysis obtained by the strongest research design (the "true experimental pretest-posttest control group design") we can state that in the short term the hypotheses with respect to the Main Idea program are largely confirmed. This means that in the short term the main idea task can be positively influenced by means of the program and that this program has no positive transfer-effect on tests measuring other tasks of reading comprehension or on tests for reading comprehension in general. From this we can deduce that the main idea task is an empirical distinctive aspect of reading comprehension.

Finally the analyses showed that the hypotheses related to the Intention program were not confirmed. For this reason we can not answer the question of the distinctiveness of the intention task and of the effect of this program on tests which in general measure reading comprehension. The results of an analysis carried out in the descriptive section of our study suggests that the intention task does form a specific aspect of reading comprehension.

On the basis of the results of the experimental section of our research no definitive answer can be given with regard to the effect of task directed instruction in relation to reading comprehension. However, a tentative recommendation seems to be permitted. This does not mean that reading comprehension in the elementary school should be completely absorbed into task directed instruction. Besides the development of strategic or procedural knowledge, adequate attention should be paid to the development of general or structured knowledge (schema-knowledge). An explanation for the very positive effect of the Reasoning program and the moderate and not positive effect of the Main Idea and Intention program respectively can probably be traced back to the weak role played by structured knowledge in the reasoning task and the relatively important role it plays in the intention task and the main idea task. The correctness of this explanation should of course be tested.

- Aarnoutse, C., Akerboom, M., Buitenhuis, A., & H. Lucas, *Enkele aspecten van lezen en leesonderwijs*. Deelrapport 1. Nijmegen: Instituut voor Onderwijskunde, 1975.
- Aarnoutse, C., Buitenhuis, A., Lucas, H., & M. Poos-Akerboom, *Doelstellingen van leesonderwijs*. Deelrapport 2. Nijmegen: Instituut voor Onderwijskunde, 1976.
- Aarnoutse, C., Buitenhuis, A., Koopman, W., Mommers, M., & M. Poos-Akerboom, *Een onderzoek naar de mate waarin een aantal doelstellingen van begrip- pend lezen wordt bereikt*. Verslag van een literatuurstudie en een voor- onderzoek. Deelrapport 3. Nijmegen: Instituut voor Onderwijskunde, 1977.
- Aarnoutse, C., Buitenhuis, A., Koopman, W., Poos-Akerboom, M., & G. Seegers, *Begrijpend lezen in het vierde leerjaar van de basisschool*. ontwikke- ling van meetinstrumenten en programma's. Deelrapport 4. Nijmegen: In- stituut voor Onderwijskunde, 1979.
- Aarnoutse, C.A.J., Begrijpend lezen in het vierde leerjaar van de basis- school. In: M.J.C. Mommers & B.W.G.M. Smits (Eds.), *Lees- Taalonder- wijs in de basisschool. Bijdragen tot de Onderwijseresearchdagen 1979*. S.V.O.-reeks no. 24. 's-Gravenhage: Staatsuitgeverij, 1979, pp. 82-105.
- Aarnoutse, C., Buitenhuis, A., Koopman, W., & G. Seegers, *Begrijpend lezen in het vierde leerjaar van de basisschool*. Deelrapport 5, Nijmegen: Instituut voor Onderwijskunde, 1980.
- Abelson, R.P., Concepts for representing mundane reality in plans. In: D.G. Bobrow & A.M. Collins (Eds.), *Representation and understanding: Studies in cognitive science*. New York: Academic Press, 1975, pp. 273-309.
- Anderson, R.C., How to construct achievement tests to assess comprehension. *Review of Educational Research*, 1972, 42(2), 145-170.
- Anderson, R.C., The notion of schemata and the educational enterprise: ge- neral discussion on the conference. In: R.C. Anderson, R.J. Spiro & W.E. Montague (Eds.), *Schooling and the acquisition of knowledge*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977, pp. 415-431.
- Anderson, R.C., Schema-directed processes in language comprehension. In: A.M. Lesgold, J.W. Pellegrino, S.D. Fokkema, & R. Glaser (Eds.), *Cogni- tive psychology and instruction*. New York: Plenum Press, 1978, pp. 67-82.
- Anderson, R.C., & J.W. Pichert, Recall of previously unrecallable informa- tion following a shift in perspective. *Journal of Verbal Learning and Verbal Behavior*, 1978, 17, 1-12.
- Anderson, R.C., Spiro, R.J., & M.C. Anderson, *Schemata as scaffolding for the representation of information in connected discourse*. Technical Report no. 24. Center for the Study of Reading. University of Illinois at Urbana-Champaign, 1977.
- Andrich, D., & J.R. Godfrey, Hierarchies in the skills of Davis' Reading Comprehension Test, Form D: an empirical investigation using a latent trait model. *Reading Research Quarterly*, 1978-1979, 14(2), 182-200.
- Appelhof, P.N., *Begeleide onderwijsvernieuwing. Evaluatie van een curriculum- innovatie gericht op differentiatie van het aanvankelijk leesonderwijs*. Tilburg: Zwijzen, 1979.
- Athey, I., Comments on design for developing comprehension skills. In: J.T. Guthrie (Ed.), *Cognition, curriculum and comprehension*. Newark, Del.: International Reading Association, 1977, pp. 233-241.

- Bakker, J., *Stilleestests, bestemd voor het derde, vierde en zesde leerjaar van de basisschool*. Nijmegen: Berkhout Testmateriaal, 1972.
- Barclay, J.R., The role of comprehension in remembering sentences. *Cognitive Psychology*, 1973, 4, 229-254.
- Barrett, T.C., *Taxonomy of cognitive and affective dimensions of reading comprehension*. Madison, Wisc.: Wisconsin Research and Development Center for Cognitive Learning, 1968.
- Bartlett, F.C., *Remembering: a study in experimental and social psychology*. Cambridge: University Press, 1932.
- Berg, R.M. van den, & H.G. te Lintelo, *Individualisering van het leesonderwijs*. Map 3. Analyse inzichtelijk lezen. 's-Hertogenbosch: Katholiek Pedagogisch Centrum, 1974.
- Beukhof, G., et al., *Handleiding bij de leergang en het introductieprogramma I. Project Stuba*. Utrecht: Vakgroep Onderwijskunde, 1979.
- Bickley, A.C., Ellington, B.J., & R.T. Bickley, The cloze procedure: a conspectus. *Journal of Reading Behavior*, 1970, 2, 232-249.
- Blaricom, G. van, & S. White, *Identifying passage meaning skills*. A paper presented for the symposium: Identification of Comprehension Skills at the 1975 Annual Meeting of the National Reading Conference. St. Petersburg, 1975.
- Blaricom, G. van, & S. White, *The effect of passage organization on main idea comprehension at three response levels*. Madison, Wisc.: Wisconsin Research and Development Center, 1976. ED 123573.
- Bloom, B.S. (Ed.), *Taxonomy of educational objectives. The classification of educational goals. Handbook I: Cognitive domain*. New York: David McKay Company, 1956.
- Bloom, B.S., Hastings, J.T., & G.F. Madaus, *Handbook of formative and summative evaluation of student learning*. New York: McGraw-Hill Book Company, 1971.
- Bobrow, D.G., Dimensions of representations. In: D.G. Bobrow & A.M. Collins (Eds.), *Representation and understanding: Studies in cognitive science*. New York: Academic Press, 1975, pp. 2-34.
- Bobrow, D.G., & A.M. Collins (Eds.), *Representation and understanding: Studies in cognitive science*. New York: Academic Press, 1975.
- Bol, E., Het meten van leesvaardigheid door middel van vragen. *Tijdschrift voor Onderwijsresearch*, 1978, 3(5), 201-215.
- Bol, E., Onderzoeksoptzet voor begrijpend lezen op de basisschool (derde tot en met zesde leerjaar). In: M.J.C. Mommers & B.W.G.M. Smits (Eds.), *Lees- Taalonderwijs in de basisschool. Bijdragen tot de Onderwijsresearchdagen 1979*. S.V.O.-reeks no. 24. 's-Gravenhage: Staatsuitgeverij, 1979, pp. 106-127.
- Bol, E., *Task structure analysis of informative texts with regard to the teaching of reading in the primary school*. Utrecht: Psychologisch Laboratorium, 1980.
- Bol, E., Köllen, E., & J. Leistra, *Begrijpend lezen. Concept draaiboek 3e klas*. S.V.O.-project 0275. Utrecht: Psychologisch Laboratorium, 1979.
- Bond, G.L., & R. Dijkstra, The cooperative research program in first grade reading. *Reading Research Quarterly*, 1967, 2, 4-142.
- Bond, G.L., & E. Wagner, *Teaching the child to read*. New York: The Macmillan Company, 1960.
- Bormuth, J.R., Comparable cloze and multiple-choice comprehension test scores. *Journal of Reading*, 1967, 10, 291-299.

- Bormuth, J R., Factor validity of cloze tests as measure of reading comprehension ability *Reading Research Quarterly*, 1969, 3, 358-367.
- Bormuth, J.R., *On the theory of achievement test items. With an appendix by P. Menzel: On the linguistic bases of the theory of writing items.* Chicago University of Chicago Press, 1970
- Bormuth, J R., Reading literacy: its definition and assessment. *Reading Research Quarterly*, 1973-1974, 9, 7-66
- Bormuth, J R , *Standardized norm-referenced tests of comprehension.* Draft. 1975 (zonder plaatsaanduiding).
- Bossers, A., *Van "Stillezen en zelf doen" naar "Stillezen en er wat mee doen".* Verslag van een onderzoek- en ontwikkelingsproject 1970-1974. Amsterdam Kohnstamm Instituut voor Onderwijsresearch van de Universiteit van Amsterdam, 1974.
- Bossers, A., *Stillezen en er wat mee doen.* Deel I voor klas 4. Handleiding. Groningen: Wolters-Noordhoff, 1975.
- Bower, G.H , Experiments on story understanding and recall. *Quarterly Journal of Experimental Psychology*, 1976, 28, 511-534.
- Bransford, J.D., Barclay, J.D., & J.J Franks, Semantic memory: a constructive versus interpretive approach, *Cognitive Psychology*, 1972, 2, 193-209.
- Bransford, J.D., & M.K. Johnson, Considerations of some problems of comprehension. In: W.G. Chase (Ed.), *Visual information processing.* New York: Academic Press, 1973, pp. 383-438.
- Bransford, J.D., & N.S. McCarrell, A sketch of a cognitive approach to comprehension. Some thoughts about understanding what it means to comprehend. In: W B. Weimer & D.S. Palermo (Eds.), *Cognition and the symbolic processes.* Hillsdale, N.J.. Lawrence Erlbaum Associates, 1974, pp. 189-229.
- Brewer, W.F., Is reading a letter-by-letter process? In. J.F. Kavanagh & I.G. Mattingly (Eds.), *Language by ear and by eye. The relationships between speech and reading.* Cambridge, Mass.: The MIT Press, 1972, pp. 359-365.
- Brewer, W.F., *Memory for the pragmatic implications of sentences.* Unpublished manuscript. University of Illinois at Urbana-Champaign, 1974.
- Brouwer, R.H.M., Onderzoek naar de leesmoeilijkheid van Nederlands proza. *Pedagogische Studien*, 1963, 40, 454-465.
- Brown, A.L., & S.S. Smiley, Rating the importance of structural units of prose passages A problem of metacognitive development. *Child Development*, 1977, 48, 1-8.
- Brus, B.Th., & J.B.A.M. van Bergen, *Schriftelijke Opdrachten.* Schoolvoordringentest voor het begrijpend lezen, bestemd voor het tweede, derde en vierde leerjaar van het basisonderwijs. Nijmegen: Berkhout Testmateriaal, 1973.
- Brus, B.Th., & M.J.M. Voeten, *Eén-Minuut-Test. Vorm A en B.* Schoolvoordringentest voor de technische leesvaardigheid, bestemd voor het tweede tot en met het zesde leerjaar van het basisonderwijs. Nijmegen: Berkhout Testmateriaal, 1973.
- Buithuis, A.F., *Nadere analyse van toetsen voor begrijpend lezen. Toepassing van latente trek analyse en generaliseerbaarheidstheorie op enkele doelstellersgerichte toetsen voor het meten van aspecten van begrijpend lezen.* Nijmegen. Instituut voor Onderwijskunde, Interne Mededelingen no. 7, 1981.

- Bulte, J., et al., *Criteriumtoetsen en de hiërarchie van onderwijsdoelstellingen. Een onderzoek m.b.t. het algoritme voor vermenigvuldigen*. Nijmegen: Instituut voor Onderwijskunde, Interne Mededelingen no. 3, 1979.
- Caesar, F.B., *Handleiding: Veilig leren lezen*. Tilburg: Zwijsen, 1970.
- Calcar, C. van, Tellegen, B., & W. van Soest, *Handleiding voor de toets 'Begrijpend Lezen I'*. Amsterdam: Research Instituut voor Toegepaste Psychologie, 1970.
- Calfee, R.C., & P.A. Drum, *How the researcher can help the teacher with classroom assessment*. Draft. Stanford: z.j.
- Calfee, R.C., & P.A. Drum, *Learning to read: Theory, Research and Practice. Curriculum Inquiry*, 1978, 8(3), 183-249.
- Campbell, D.T., & J.C. Stanley, *Experimental and quasi-experimental designs for research on teaching*. In: N.L. Gage (Ed.), *Handbook of research on teaching*. Chicago: Rand McNally, 1971.
- Carroll, J.B., *Defining language comprehension: Some speculations*. In: R.O. Freedle & J.B. Carroll (Eds.), *Language comprehension and the acquisition of knowledge*. Washington D.C.: V.H. Winston, 1972, pp. 1-29.
- Carroll, C.A., *Low achievers' understanding of logical inference forms*. In: M.F. Roskopf (Ed.), *Children's mathematical concepts. Six Piagetian studies in mathematics education*. New York: Teachers College Press, 1975, pp. 173-208.
- Carver, R.P., *Analysis of 'chunked' test items as measures of reading and listening comprehension*, *Journal of Educational Measurement*, 1970, 7, 141-150.
- Centraal Bureau voor de Statistiek, *Typologie van de Nederlandse gemeenten naar urbanisatiegraad, 31 mei 1960*. Zeist, 1964.
- Centraal Bureau voor de Statistiek, *Bevolking der gemeenten van Nederland op 1 januari 1975*. 's-Gravenhage, 1975.
- Centraal Instituut voor Toetsontwikkeling, *Begrijpend lezen leerjaar 3, 4 en 5*. Arnhem, 1980.
- Chapman, C.A., *An analysis of three theories of the relationships among reading-comprehension skills*. Symposium on psycholinguistics and reading presented at the meeting of the International Reading Association. Kansas City, 1969.
- Chapman, C.A., *A test of a hierarchical theory of reading comprehension*. Dissertation. Chicago, Ill.: University of Chicago, 1971.
- Chester, R.D., et al., *Teacher's planning guide: Comprehension*. Madison, Wisc.: Wisconsin Research and Development Center for Cognitive Learning, 1974.
- Clark, H.H., *Inferences in comprehension*. In: D. LaBerge & S.J. Samuels (Eds.), *Basic processes in reading: Perception and comprehension*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977, pp. 243-263.
- Clymer, T., *What is 'reading'? Some current concepts*. In: R.M. Robinson (Ed.), *Innovation and change in reading instruction. Sixty-seventh Yearbook of the National Society of the Study of Education, Part II*. Chicago, Ill.: University of Chicago Press, 1968, pp. 7-29.
- Cohen, J., *Statistical power analysis for the behavioral sciences*. New York: Academic Press, 1969.
- Collins, A., & E.E. Smith, *Teaching the process of reading comprehension*. Technical Report no. 182. Center for the Study of Reading. University of Illinois at Urbana-Champaign, 1980.

- Cook, Th.D., & D.T. Campbell, The design and conduct of quasi-experiments and true experiments in field settings. In: M.D. Dunnnette (Ed.), *Handbook of industrial and organizational psychology*. Chicago: Rand McNally, 1976, pp. 223-319.
- Daandels, W., *Een onderzoek naar het onderwijs van het begrijpend lezen op Nederlandse basisscholen*. Utrecht: Psychologisch Laboratorium, 1974.
- Dahl, Th.A., *The measurement of congruence between learning objectives and test items*. Dissertation. Los Angeles: University of California, 1971.
- Danhof, W., Smelt, L., & A. van der Wissel, De relatie tussen technisch en begrijpend lezen op twee leeftijdsniveaus. *Pedagogische Studiën*, 1978, 55, 462-471.
- Davis, F.B., Fundamental factors of comprehension in reading. *Psychometrika*, 1944, 9(3), 185-197.
- Davis, F.B., Research in comprehension in reading. *Reading Research Quarterly*, 1968, 3(4), 499-545.
- Davis, F.B., *The literature of research in reading with emphasis on models*. Final report. New Brunswick, N.J.: University of Pennsylvania, 1971.
- Davis, F.B., Psychometric research on comprehension in reading. *Reading Research Quarterly*, 1972, 7(4), 628-678.
- Day, J.D., *Teaching summarization skills: A comparison of training methods*. Unpublished doctoral dissertation. University of Illinois, 1980.
- Deen, N., *Een halve eeuw onderwijsresearch in Nederland. Het Nutsseminarium voor Pedagogiek aan de Universiteit van Amsterdam 1919-1969*. Groningen: Wolters-Noordhoff, 1969.
- Derrick, C., *Three aspects of reading comprehension as measured by tests of different growth*. Unpublished doctoral dissertation. University of Chicago, 1953.
- Dijk, T.A. van, Semantic macro-structures and knowledge frames in discourse comprehension. In: M.A. Just & P.A. Carpenter (Eds.), *Cognitive processes in comprehension*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977, pp. 3-32.
- Dijk, T.A. van, *Tekstwetenschap. Een interdisciplinaire inleiding*. Utrecht: Het Spectrum, 1978.
- Dongen, D. van, & H. van Leent, *Preventie van leesmoelijkheden. Opzet van het eerste deel van het exploratieve onderzoek. Deelrapport 4*. Nijmegen: Instituut voor Onderwijskunde, 1981.
- Drenth, P.J.D., *De psychologische test. Een inleiding in de theorie van de psychologische test en zijn toepassingen*. Deventer: Van Loghum Slaterus, 1972⁴.
- Drop, W., & J.H.L. de Vries, *Taalbeheersing. Handboek voor taalhantering*. Groningen: Wolters-Noordhoff, 1977.
- Dumont, J.J., *Leerstoornissen. Deel 1 en 2*. Rotterdam: Lemniscaat, 1971.
- Durkin, D., *Children who read early*. Columbia University: Teachers College Press, 1966.
- Durkin, D., *Comprehension instruction - where are you?* Reading Education Report no. 1. Center for the Study of Reading: University of Illinois at Urbana-Champaign, 1977.
- Durkin, D., What classroom observations reveal about reading comprehension instruction. *Reading Research Quarterly*, 1978-1979, 14(4), 481-533.

- Ebel, R.L., Educational tests: valid? biased? useful? *Phi Delta Kappan*, 1975, 57(2), 83-88.
- Eeuwijk-Rijnen, A.J.M.L.A. van, et al., *Met inzicht leren lezen*. Tilburg: Zwijsen, 1973.
- Elkonin, D.B., The psychology of mastering the elements of reading. In: B. Simon & J. Simon (Eds.), *Educational psychology in the USSR*. London, 1963. (Geciteerd door Otto, 1971).
- Ennis, R.H., An alternative to Piaget's conceptualization of logical competence. *Child Development*, 1976, 47, 903-919.
- Ennis, R.H., & D.H. Paulus, *Critical thinking readiness in grades 1-12. Phase 1: Deductive reasoning in adolescence*. Ithaca, N.Y.: Cornell Critical Thinking Project, 1965. ED 003818.
- Farr, R.C., The convergent and discriminant validity of several upper level reading tests. In: G.B. Schick & M.M. May (Eds.), *Multidisciplinary aspects of college-adult reading*. Yearbook of the National Reading Conference, 1968, no. 17, pp. 181-191.
- Farr, R.C., *Reading: what can be measured?* Newark, Del.: International Reading Association, 1969.
- Farr, R.C., *Measuring reading comprehension: An historical perspective*. Paper presented at the National Reading Conference. St. Peterburgh (Fla), 1970. ED 050890.
- Farr, R.C., & N. Anastasiow, *Tests of reading readiness and achievement: a review and evaluation*. Newark, Del.: International Reading Association, 1971².
- Finn, J.D., *Multivariate: Univariate and multivariate analysis of variance, covariance and regression*. Ann Arbor, Mich.: National Educational Resources Inc., 1972.
- Finn, J.D., *A general model for multivariate analysis*. New York: Holt, Rinehart and Winston, 1974.
- Frederiksen, C.H., Effects of task-induced cognitive operations on comprehension and memory processes. In: R.O. Freedle & J.B. Carroll (Eds.), *Language comprehension and the acquisition of knowledge*. Washington, D.C.: V.H. Winston, 1972, pp. 211-245.
- Frederiksen, C.H., Discourse comprehension and early reading. In: L.B. Resnick & P.A. Weaver (Eds.), *Theory and practice of early reading*. Volume 1. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1979, pp. 155-186.
- Gagné, R.M., *Essentials of learning for instruction*. New York: Holt, Rinehart and Winston, 1974.
- Gagné, R.M., Schooling and the relevance of research: general discussion. In: R.C. Anderson, R.J. Spiro & W.E. Montague (Eds.), *Schooling and the acquisition of knowledge*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977, pp. 405-413 (a).
- Gagné, R.M., *The conditions of learning*. New York: Holt, Rinehart and Winston, 1977³ (b).
- Gates, A.I., *The improvement of reading*. New York: The Macmillan Company, 1947.
- Geyer, J.J., Comprehensive and partial models related to the reading process. *Reading Research Quarterly*, 1972, 7(4), 541-587.
- Gibson, E.J., & H. Levin, *The psychology of reading*. Cambridge, Mass.: The MIT Press, 1975.

- Glaser, R., & A.J. Nitko, Measurement in learning and instruction. In: R.L. Thorndike (Ed.), *Educational measurement*. Washington D.C.: American Council on Education, 1971.
- Glaser, R., Pellegrino, J.W., & A.M. Lesgold, Some directions for a cognitive psychology of instruction. In: A.M. Lesgold, J.W. Pellegrino, S.D. Fokkema & R. Glaser (Eds.), *Cognitive psychology and instruction*. New York: Plenum Press, 1978, pp. 495-517.
- Goetz, E.T., *Inferences in the comprehension of and memory for text*. Technical Report no. 49. Center for the Study of Reading. University of Illinois at Urbana-Champaign, 1977.
- Golinkoff, R.M., & R.R. Rosinsky, Decoding, semantic processing, and reading comprehension skill. *Child Development*, 1976, 47, 252-258.
- Goodlad, J.I., Klein, M.F., & Associates, *Behind the classroom door*. Worthington, Ohio: Charles A. Jones Publishing Company, 1970.
- Goodlad, J.I., & Associates, *Curriculum Inquiry. The study of curriculum practice*. New York: McGraw-Hill Book Company, 1979.
- Goodman, K.S., The psycholinguistic nature of the reading process. In: K.S. Goodman (Ed.), *The psycholinguistic nature of the reading process*. Detroit, Mich.: Wayne State University Press, 1968, pp. 15-27.
- Goodman, K.S., Behind the eye: what happens in reading. In: K.S. Goodman & O.S. Niles (Eds.), *Reading: process and program*. Urbana, Ill.: NCTE, 1970, pp. 3-39.
- Goodman, K.S., Reading: the key is in the children's language. *The Reading Teacher*, 1972, 505-508.
- Goodman, K.S., The psycholinguistic nature of the reading process. In: K.S. Goodman (Ed.), *The psycholinguistic nature of the reading process*. Detroit, Mich.: Wayne State University Press, 1973², pp. 13-26.
- Goodman, K.S., Reading: a psycholinguistic guessing game. In: H. Singer & R.B. Ruddell (Eds.), *Theoretical models and processes in reading*. Newark, Del.: International Reading Association, 1976, pp. 497-508.
- Goodman, K.S., Strategies for increasing comprehension in reading. In: Scott, *Foresman Monograph*. Palo Alto, Cal.: Scott, Foresman and Company, z.j.
- Gordon, C.J., *The effects of instruction in metacomprehension and inferencing on children's comprehension abilities*. Unpublished dissertation. University of Minnesota, 1979.
- Gough, P.B., One second of reading. In: J.F. Kavanagh & I.G. Mattingly (Eds.), *Language by ear and by eye. The relationships between speech and reading*. Cambridge, Mass.: The MIT Press, 1972.
- Gray, W.S., The major aspects of reading. In: H.M. Robinson (Ed.), *Sequential development of reading abilities*. Proceedings of the annual conference on reading held at the University of Chicago, 1960. Supplementary Educational Monographs no. 90. Chicago: University of Chicago Press, 1960, pp. 8-24.
- Greene, F., Radio reading. In: C. Pennoch (Ed.), *Reading comprehension at four linguistic levels*. Newark, Del.: International Reading Association, 1979, pp. 104-107.
- Guthrie, J.T., Models of reading and reading disability. *Journal of Educational Psychology*, 1973, 65, 9-18.
- Guthrie, J.T. (Ed.), *Cognition, curriculum and comprehension*. Newark, Del.: International Reading Association, 1977.

- Hamaker, C., Over "representatie". Slotbeschouwing van de ORD'80 zitting gewijd aan tekstbestudering. In: C. Boonman & J. Zwarts (Eds.), *Tekstbestudering. Bijdragen tot de Onderwijsresearchdagen 1980*. S.V.O.-reeks no. 35. 's-Gravenhage: Staatsuitgeverij, 1980, pp. 202-217.
- Hansen, J., The effects of inference training and practice on young children's reading comprehension. *Reading Research Quarterly*, 1981, 16(3), 391-417.
- Hansen, J., & P.D. Pearson, *The effects of inference training and practice on young children's comprehension*. Technical Report no. 166. Center for the Study of Reading. University of Illinois at Urbana-Champaign, 1980.
- Harris, A.J., & E.R. Sipay, *How to increase reading ability*. New York: David McKay Company, 1977.
- Harris, C.W., Measurement of comprehension in literature. *School Review*, 1948, 56, 280-289; 332-342.
- Harris, P.L., Tekstbegrip bij jonge kinderen. In: J. de Wit, H. Bolle & J.M. van Meel (Eds.), *Psychologen over het kind*. No. 6. Informatie over de ontwikkeling van en hulpverlening aan kinderen en jeugd. Groningen: Wolters-Noordhoff, 1979, pp. 31-47.
- Harris, Th. L., Reading. In: R. Ebel (Ed.), *Encyclopedia of educational research*. New York: MacMillan, 1969, pp. 1069-1104.
- Hatcher, T.C., *The development of comprehension skills in selected basal readers*. Dissertation. Ohio State University, 1971.
- Hauwermeiren, P. van, *Het leesbaarheidsonderzoek*. Groningen: Tjeenk Willink, 1975.
- Haviland, S.E., & H.H. Clark, What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior*, 1974, 13, 512-521.
- Hill, S.A., *A study of logical abilities in children*. Unpublished doctoral dissertation. Stanford University, 1961.
- Hintzman, D.L., Comments on the papers by Hunt and by Rumelhart and Norman. In: J.W. Cotton & R.L. Klatzky (Eds.), *Semantic factors in cognition*. Hillsdale N.J.: Lawrence Erlbaum Associates, 1978, pp. 55-60.
- Hoeve, A.H. van der, Kohnstamm, Ph., & G. van Veen, *De lagere school en het stil-lezen*. (Een onderzoek inzake het zelfstandig verwerken van leesstof door leerlingen der hoogste klassen van lagere scholen in de gemeente Utrecht). Mededelingen van het Nutsseminarium voor Paedagogiek aan de Universiteit van Amsterdam. No. 23. Groningen: Wolters, 1933.
- Hoeve, A.H. van der, Kohnstamm, Ph., & G. van Veen, *Stil-lees-stof als denk-materiaal en denk-maatstaf. Een onderzoek aangaande de vorming en toetsing van "Theoretische Intelligentie"*. Mededelingen van het Nutsseminarium voor Paedagogiek aan de Universiteit van Amsterdam. No. 24. Groningen: Wolters, 1935.
- Huey, E.B., *The psychology and pedagogy of reading*. Reprint. Cambridge, Mass.: The MIT Press, 1968 [1908].
- Hunt, L.C. jr., The effects of self-selection, interest, and motivation upon independent, instructional and frustration levels. *The Reading Teacher*, 1970, 24(2), 146-151.
- Johnson, M.K., Bransford, J.D., & S.K. Solomon, Memory for tacit implications of sentences. *Journal of Experimental Psychology*, 1973, 98, 203-205.

- Jolly, H.B. jr., Determining main ideas: a basic study skill. In: L.E. Hafner, *Improving reading in middle and secondary schools*. New York: Macmilland Publishing Co., 1974, pp. 162-172.
- Jongsma, E.R., *The cloze procedure: a survey of research*. Occasional papers in reading. Bloomington, Ind., 1970. ED 050893.
- Kamm, K., A five year study of the effects of a skill-centered approach to the teaching of reading. *The Journal of Educational Research*, 1978, 72(2), 104-112.
- Kidder, S.J., & R.T. Schuder, *Measurement and productivity in school reading programs*. Main report. Albany, N.Y.: The University of the State of New York, The State Education Department, Division of Research, 1975.
- Kintsch, W., *The representation of meaning in memory*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1974.
- Kintsch, W., On comprehending stories. In: M.A. Just & P.A. Carpenter (Eds.), *Cognitive processes in comprehension*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977, pp. 33-62 (a).
- Kintsch, W., Reading comprehension as a function of text structure. In: A.S. Reber & D.L. Scarborough (Eds.), *Toward a psychology of reading*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977, pp. 227-256 (b).
- Kintsch, W., & T.A. van Dijk, Toward a model of text comprehension and production. *Psychological Review*, 1978, 85(5), 363-394.
- Kintsch, W., Kozminsky, E., Streby, W.J., McKoon, G., & J.M. Keenan, Comprehension and recall of text as a function of content variables. *Journal of Verbal Learning and Verbal Behavior*, 1975, 14, 196-214.
- Klerk, L.F.W. de, *Inleiding in de onderwijspsychologie*. Deventer: Van Loghum Slaterus, 1979.
- Kohnstamm, Ph., *De formele logica en het kinderlijke denken*. Mededelingen van het Nutsseminarium voor Paedagogiek aan de Universiteit van Amsterdam. No. 26. Groningen: Wolters, 1934.
- Kohnstamm, Ph., *Keur uit het didactisch werk van prof.dr. Ph. Kohnstamm*. Tweede druk. Met een korte schets uit zijn leven en ontwikkelingsgang door prof.dr. M.J. Langeveld. Groningen: Wolters, 1952.
- Kok, W., Boonman, C., & G. Beukhof, Studievaardigheden voor de basisschool: leren omgaan met teksten I, II. *Pedagogische Studiën*, 1980, 57(10, 11/12), 417-432, 493-503.
- LaBerge, D., & S.J. Samuels, Toward a theory of automatic information processing in reading. *Cognitive Psychology*, 1974, 6, 293-323.
- Lefreuve, C.A., *Linguistics and the teaching of reading*. New York: McGraw-Hill, 1964.
- Leistra, J., *De hoofdgedachte van een tekst (een schets)*. Ongepubliceerd paper. Amsterdam: Vakgroep Didactiek, 1979.
- Levelt, W.J.M., & G.A.M. Kempen, Taal. In: J.A. Michon, E.G.J. Eijkman & L.F.W. de Klerk (Eds.), *Handboek der psychonomie*. Deventer: Van Loghum Slaterus, 1976, pp. 492-523.
- Lord, F.M., & M.R. Novick, *Statistical theories of mental test scores*. Reading, Mass.: Addison-Wesley Publishing Company, 1974.
- Lowerre, G.F., & J.M. Scandura, Conceptually based development and evaluation of individualized materials for critical reading based on logical inference. *Reading Research Quarterly*, 1973-1974, 9(2), 186-205.

- Malmquist, E., & B.Th. Brus, *Lezen leren, lezend leren. Beschouwingen en aanwijzingen met betrekking tot het leesonderwijs in al zijn facetten*. Tilburg: Zwijsen, 1974.
- Mandler, J.M., & N.S. Johnson, Remembrance of things parsed: story structure and recall. *Cognitive Psychology*, 1977, 9, 111-151.
- Manoliu-Dabija, A., & A.C. Schipper, Het maken van afleidingen tijdens het lezen en enige beschouwingen over het probleemoplossingsmodel voor het leren. In: C. Boonman & J. Zwarts (Eds.), *Tekstbestudering. Bijdragen tot de Onderwijsresearchdagen 1980*. S.V.O.-reeks no. 35. 's-Gravenhage: Staatsuitgeverij, 1980, pp. 74-109.
- Manzo, A.V., The request procedure. *Journal of Reading*, 1969, 13(2), 123-126.
- McConkie, G.W., & K. Rayner, Identifying the span of the effective stimulus in reading: literature review and theories of reading. In: H. Singer & R.B. Ruddell (Eds.), *Theoretical models and processes of reading*. Newark, Del.: International Reading Association, 1976², pp. 137-162.
- Miller, G.A., & P.N. Johnson-Laird, *Language and perception*. Cambridge, Mass.: Harvard University Press, 1976.
- Minsky, M.A., A framework for representing knowledge. In: P.H. Winston (Ed.), *The psychology of computer vision*. New York: McGraw-Hill, 1975, pp. 211-277.
- Mommers, M.J.C., & B.W.G.M. Smits, *Toetsen voor inzichtelijk rekenen. Verantwoording en handleiding*. Tilburg: Zwijsen, 1973.
- Nuy, H.J.P., *Constructie en gedeeltelijke evaluatie van een leergang begrijpend lezen voor de tweede klas basisschool*. Rotterdam: Project Onderwijs en Sociaal Milieu, 1980.
- Olshavsky, J.E., Reading as problem solving: an investigation of strategies. *Reading Research Quarterly*, 1976-1977, 12(4), 654-675.
- Oostendorp, H. van, & M. den Uyl, Regulatie van inferentie en integratie bij het lezen van teksten. In: C. Boonman & J. Zwarts (Eds.), *Tekstbestudering. Bijdragen tot de Onderwijsresearchdagen 1980*. S.V.O.-reeks no. 35. 's-Gravenhage: Staatsuitgeverij, 1980. pp. 134-161.
- O'Reilly, R.P., Schuder, R.T., & S.J. Kidder, *Validation of a multiple choice cloze test of literal comprehension: Summary Report*. Albany, N.Y.: The University of the State of New York, The State Education Department, Division of Research, 1976.
- O'Reilly, R.P., & R.E. Streeter, Report on the development and validation of a system for measuring literal comprehension in a multiple-choice cloze format: Preliminary factor analytic results. *Journal of Reading Behavior*, 1977, 9(1), 45-69.
- Otto, W., Evaluating instruments for assessing needs and growth in reading. In: W.H. MacGinitie (Ed.), *Assessment problems in reading*. Newark, Del.: International Reading Association, 1973, pp. 14-20.
- Otto, W., Design for developing comprehension skills. In: J.T. Guthrie (Ed.), *Cognition, curriculum and comprehension*. Newark, Del.: International Reading Association, 1977, pp. 193-232.
- Otto, W., & Th.C. Barrett, *Two studies of children's ability to formulate and state a literal main idea in reading*. Technical Report no. 57. Madison, Wisc.: Wisconsin Research and Development Center for Cognitive Learning, 1968. ED 024543.

- Otto, W., Barrett, Th.C., & K. Koenke, *The assessment of children's state-ments of the main idea in reading*. Madison, Wisc.: Wisconsin Research and Development Center for Cognitive Learning, 1968. ED 019197.
- Palland, B.G., *Een intelligentie-onderzoek op een tiental Amsterdamse scholen*. Mededelingen van het Nutsseminarium voor Paedagogiek aan de Universiteit van Amsterdam. No. 25. Groningen: Wolters, 1935.
- Pavlak, S.A., *Reading comprehension. A critical analysis of selected factors affecting comprehension*. Dissertation. University of Pittsburgh, 1973.
- Perfetti, C.A., Language comprehension and fast decoding: some psycholin-guistic prerequisites for skilled reading comprehension. In: J.T. Guthrie (Ed.), *Cognition, curriculum, and comprehension*. Newark, Del.: International Reading Association, 1977, pp. 20-41.
- Perfetti, C.A., & A.M. Lesgold, Discourse comprehension and sources of in-dividual differences. In: M.A. Just & P.A. Carpenter (Eds.), *Cognitive processes in comprehension*. Hillsdale, N.J.: Lawrence Erlbaum Associ-ates, 1977, pp. 141-183.
- Piaget, J., *The psychology of intelligence*. New York: Harcourt, Brace and World, 1963.
- Pichert, J.W., & R.C. Anderson, Taking different perspectives on a story. *Journal of Educational Psychology*, 1977, 69(4), 309-315.
- Popham, J., *Educational evaluation*. Englewood Cliffs, N.J.: Prentice Hall, 1975.
- Popham, J., *Criterion-referenced measurement*. Englewood Cliffs, N.J.: Prentice Hall, 1978.
- Potter, T.C., A taxonomy of cloze research, part I: *Readability and rea-ding comprehension*. Technical Report no. 1. Inglewood, Cal.: South-west Regional Laboratory for Educational Research and Development, 1968.
- Potts, G.R., Information processing strategies used in the encoding of li-near orderings. *Journal of Verbal Learning and Verbal Behavior*, 1972, 11, 727-740.
- Rankin, E.F., Cloze procedure - a survey of research. In: E.S. Thurston & L.E. Hafner (Eds.), *The philosophical and sociological bases of rea-ding*. Fourteenth yearbook of the National Reading Conference. Milwau-kee: National Reading Conference. 1965, pp. 133-150.
- Rankin, E.F., GAP reading comprehension test. In: O.K. Buros (Ed.), *The seventh mental measurements yearbook*. Highland Park, N.J.: Gryphon Press, 1972, pp. 1078-1080.
- Rankin, E.F., The cloze procedure revisited. In: P.L. Nacke (Ed.), *Inter-action: research and practice for college-adult reading*. Twentythird Yearbook of the National Reading Conference, Clemson, South Carolina, 1974, pp. 1-8.
- Raven, J.C., *Guide to the standard progressive matrices. Sets A, B, C, D and E*. London: Lewis, 1958.
- Raven, J.C., *Guide to the standard progressive matrices. Sets A, B, C, D and E*. London: Lewis, 1960 (Reprinted with additional Bibliography).
- Reitsma, P., De verwerving van schriftelijke taal. In: J. de Wit, H. Bolle & J.M. van Meel (Eds.), *Psychologen over het kind. Informatie over de ontwikkeling van en hulpverlening aan kinderen en jeugd*. Groningen: Wolters-Noordhoff, 1979, pp. 87-120.

- Resnick, L.B., & I.L. Beck, Designing instruction in reading: interaction of theory and practice. In: J.T. Guthrie (Ed.), *Aspects of reading acquisition*. Baltimore: John Hopkins University Press, 1976, pp. 180-204.
- Rystrom, R., The Carver-Darby chunked reading test. In: O.K. Buros (Ed.), *The seventh mental measurements yearbook*. Highland Park, N.J.: Gryphon Press, 1972, pp. 1073-1074.
- Roberge, J.J., A study of children's ability to reason with basic principles of deductive reasoning. *American Educational Research Journal*, 1970, 7, 583-596.
- Robinson, H.M., The major aspects of reading. In: H.A. Robinson (Ed.), *Reading seventy-five years of progress*. Supplementary Educational Monographs no. 96. Chicago: University of Chicago Press, 1966, pp. 22-32.
- Rosenshine, B.V., Skill hierarchies in reading comprehension. In: R.J. Spiro, B.C. Bruce & W.F. Brewer (Eds.), *Theoretical issues in reading comprehension*. Perspectives from cognitive psychology, linguistics, artificial intelligence, and education. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1980, pp. 535-554.
- Rumelhart, D.E., Toward an interactive model of reading. In: S. Dornic (Ed.), *Attention and performance*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977, pp. 573-603 (a).
- Rumelhart, D.E., Understanding and summarizing brief stories. In: D. LaBerge & S.J. Samuels (Eds.), *Basic processes in reading: perception and comprehension*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977, pp. 265-303 (b).
- Rumelhart, D.E., & D.A. Norman, Accretion, tuning and restructuring: three modes of learning. In: J.W. Cotton & R.L. Klatzky (Eds.), *Semantic factors in cognition*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1978, pp. 37-60.
- Rumelhart, D.E., & A. Ortony, The representation of knowledge in memory. In: R.C. Anderson, R.J. Spiro & W.E. Montague (Eds.), *Schooling and the acquisition of knowledge*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977, pp. 99-135.
- Samuels, S.J., Hierarchical subskills in the reading acquisition process. In: J.T. Guthrie (Ed.), *Aspects of reading acquisition*. Baltimore: John Hopkins University Press, 1976, pp. 162-179.
- Schank, R.C., The role of memory in language processing. In: C.N. Cofer (Ed.), *The structure of human memory*. San Francisco: W.H. Freeman and Company, 1976, pp. 162-189.
- Schank, R.C., & R.P. Abelson, *Scripts, plans, goals and understanding. An inquiry into human knowledge structures*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.
- Schoot, F. van der, et al., *Correct leren lezen*. Tilburg: Zwijsen, 1973.
- Singer, H., *IQ is and is not related to reading*. Invitational paper read at the Annual Convention of the International Reading Association. Denver, 1973. ED 088004.
- Singer, H., & R.B. Ruddell (Eds.), *Theoretical models and processes of reading*. Newark, Del.: International Reading Association, 1976².
- Singer, M., Context inferences in the comprehension of sentences. *Canadian Journal of Psychology*, 1976, 30, 39-46.

- Singer, M., *The temporal locus of inference in language comprehension*. Unpublished manuscript. University of Manitoba, 1977.
- Smith, F., *Psycholinguistics and reading*. New York: Holt, Rinehart and Winston, 1973.
- Smith, F., *Understanding reading. A psycholinguistic analysis of reading and learning to read*. New York: Holt, Rinehart and Winston, 1978.
- Spache, G.D., *The teaching of reading. Methods and results: an overview*. Bloomington, Ind.: Phi Delta Kappa, 1972.
- Spache, G.D., & E.B. Spache, *Reading in the elementary school*. Boston: Allyn and Bacon, 1973³.
- Spearrit, D., Identification of subskills of reading comprehension by maximum likelihood factor analysis. *Reading Research Quarterly*, 1972, 8(1), 92-111.
- Spiro, R.J., Remembering information from text: the "state of schema" approach. In: R.C. Anderson, R.J. Spiro & W.E. Montague (Eds.), *Schooling and the acquisition of knowledge*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977, pp. 137-165.
- Stanovich, K.E., Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly*, 1980, 16(1), 32-71.
- Stauffer, R.G., *Directing reading maturity as a cognitive process*. New York: Harper and Row, 1969.
- Stauffer, R.G., *The language-experience approach to the teaching of reading*. New York: Harper and Row, 1970.
- Stijnen, P.J.J., *Woordenschattest*. Bestemd voor het derde tot en met het zesde leerjaar van de basisschool. Nijmegen: Berkhout Testmateriaal, 1975.
- Sulin, R.A., & D.J. Dooling, Intrusion of a thematic idea in retention of prose. *Journal of Experimental Psychology*, 1974, 103(2), 255-262.
- Taylor, W.L., "Cloze procedure": a new tool for measuring readability. *Journalism Quarterly*, 1953, 30, 415-433.
- Taylor, W.L., Cloze readability scores as indices of individual differences in comprehension and aptitude. *Journal of Applied Psychology*, 1957, 41, 12-26.
- Terwel, J., Absoluut en relatief meten. *INFO*, 5, 1973, 31-54. Instituut voor Onderwijskunde, Groningen.
- Thomassen, A.J.W.M., & G.A.M. Kempen, Geheugen. In: J.A. Michon, E.G.J. Eijkman & L.F.W. de Klerk (Eds.), *Handboek der psychonomie*. Deventer: Van Loghum Slaterus, 1976, pp. 354-387.
- Thorndike, E.L., Reading as reasoning: a study of mistakes in paragraph reading. *Journal of Educational Psychology*, 1917, 8, 323-332. Reprint: *Reading Research Quarterly*, 1971, 6, 425-434.
- Thorndike, P.W., The role of inferences in discourse comprehension. *Journal of Verbal Learning and Verbal Behavior*, 1976, 15, 437-446.
- Thorndike, P.W., Cognitive structures in comprehension and memory of narrative discourse. *Cognitive Psychology*, 1977, 9, 77-110.
- Thorndike, R.L., *Reading as reasoning*. Address delivered to Division 15, American Psychological Association. Washington, 1971.
- Thorndike, R.L., "Reading as reasoning". *Reading Research Quarterly*, 1973-1974, 9(2), 135-147.

- Thurstone, L.L., Note on a reanalysis of Davis' reading tests. *Psychometrika*, 1946, 11, 185-188.
- Tierney, R.J., & J.W. Cunningham, *Research on teaching reading comprehension*. Technical Report no. 187. Center for the Study of Reading. University of Illinois at Urbana-Champaign, 1980.
- Toulmin, S., *The uses of argument*. London: Cambridge University Press, 1958.
- Trabasso, T., Mental operations in language comprehension. In: R.O. Freedle & J.B. Carroll (Eds.), *Language comprehension and the acquisition of knowledge*. Washington, D.C.: V.H. Winston, 1972, pp. 113-137.
- Tuinman, J.J., Determining the passage dependency of comprehension questions in 5 major tests. *Reading Research Quarterly*, 1973-1974, 9(2), 206-223.
- Tuinman, J.J., Criterion referenced measurement in a norm referenced context. In: S.J. Samuels (Ed.), *What research has to say about reading instruction*. Newark, Del.: International Reading Association, 1978, pp. 165-173.
- Uit den Boogaart, P.C. (Ed.), *Woordfrequenties in geschreven en gesproken Nederlands*. Utrecht: Oosthoek, Scheltema en Holkema, 1975.
- Vacca, R.T., A study of holistic and subskill instructional approaches to reading comprehension. *Journal of Reading*, 1980, 512-518.
- Veen, G. van, *Plaats en betekenis van het "stil lezen" in onze lagere scholen*. Mededelingen van het Nutseminarium voor Paedagogiek aan de Universiteit van Amsterdam. No. 8. Groningen: Wolters, 1929.
- Velema, E., *Over de noodzakelijkheid van een meer exacte onderwijskunde*. Mededelingen van het Centrum voor Didactiek en Onderwijsbegeleiding. no. 1. Technische Hogeschool Twente, 1966.
- Vellutino, F.R., *Dyslexia: Theory and research*. Cambridge: The MIT Press, 1979.
- Walter, R.B., *Historical overview of the cloze procedure*. Kean College of New Jersey, 1974. ED 094337.
- Wardrop, J.L., Design problems in reading research. In: R. Farr, S. Weintraub & B. Tone (Eds.), *Improving reading research*. Newark, Del.: International Reading Association, 1976, pp. 23-33.
- Warries, E. (Ed.), *Beheersingsleren een leerstrategie*. Serie Onderwijskunde, nr. 11. Groningen: Wolters-Noordhoff, 1979.
- Weaver, W.W., & A.J. Kingston, A factor analysis of the cloze procedure and other measures of reading and language ability. *The Journal of Communication*, 1963, 13, 252-261.
- Wesdorp, H., *Evaluatie-technieken voor het moedertaalonderwijs. Een inventarisatie van beoordelingsmethoden voor de stelvaardigheid, het begrijpend lezen, de spreek-, luister- en discussie-vaardigheid*. Eindrapport S.V.O.-project 0385. Amsterdam: Research Instituut voor de Toegepaste Psychologie, 1978.
- Westerlaak, J.M. van, Kropman, J.A., & J.W.M. Collaris, *Beroepenklapper*. Nijmegen: Instituut voor Toegepaste Sociologie, 1975.
- Wiegersma, S., *Leesvaardigheidstest voor het onderzoek naar mechanische leesvaardigheid*. Groningen: Tjeenk Willink, 1971.

- Wijnen, W.H.F.W., *Formative evaluation and educational testing*. Groningen: Centrum Onderzoek Wetenschappelijk Onderwijs, 1973.
- Wijnstra, J.M., Het gebruik van de cloze procedure als maat voor schriftelijke taalbeheersing. *Tijdschrift voor Onderwijsresearch*, 1977, 2, 262-269.
- Winograd, T., A framework for understanding discourse. In: M.A. Just & P.A. Carpenter (Eds.), *Cognitive processes in comprehension*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977, pp. 63-88.
- Wolf, W., King, M.L., & C.S. Huck, Teaching critical reading to elementary school children. *Reading Research Quarterly*, 1968, 3(4), 435-498.
- Wolf, W., Huck, C.S., King, M.L., et al., *Critical reading ability of elementary school children*. Final Report. Columbus, Ohio: The Ohio State University Research Foundation, 1967. ED 014407.
- Zeeuw, J. de, *Algemene psychodiagnostiek II. Testtheorie*. Amsterdam: Swets en Zeitlinger, 1978.
- Zondervan, F., Stam, P. van, & G. Gunneweg, De leesbaarheid van basisschoolteksten. Objectieve ordeningscriteria voor instructieve teksten. *De Nieuwe Taalgids*, 1976, 69, 426-445.
- Zwarts, M., Bijdragen tot de begripsvaliditeit van enkele leestoetsen m.b.t. de begrippen technisch en begrijpend lezen. Het D.A.L.-onderzoek 4. *Pedagogische Studiën*, 1977, 54, 398-405.

B I J L A G E N

Instanties en personen die medewerking verleenden aan het onderzoek.

1. Scholen, hoofden van scholen en leerkrachten.

R.-K. Basisschool "De Beundert" te Beuningen

Hoofd der School : Th.W.H. Bens

Leerkrachten : H.J. Jansen en J.A.M. Wolters.

R.-K. Basisschool "De Eikenhorst" te Nijmegen

Hoofd der School : Th.N. Hilgers

Leerkrachten : I. Hollaardt, A. Ogier en J. Smulders.

R.-K. Basisschool "Druten Centrum" te Druten

Hoofd der School : S.J. Nota

Leerkrachten : W.J.W. Heymink, J.M. van Thiel en J.M. van Uiter.

R.-K. Basisschool "Het Tolhuis" te Nijmegen

Hoofd der School : J.G. Bekker

Leerkrachten : H.A.M. Beerten en P.W.B. Verhoogt

R.-K. Basisschool "Het Vossenhol" te Groesbeek

Hoofd der School : J.A.M. van de Wijden

Leerkrachten : I. van Geelen, A. Peters en T. Voeten.

R.-K. Basisschool "Paulus" te Wijchen

Hoofd der school : A.E. Hilgers

Leerkrachten : W.A.M. van de Loop en H.L. Schipstal.

Openbare Prins Willem-Alexanderschool te Wijchen

Hoofd der School : W.A. Crum

Leerkrachten : B. van de Berg en H. Clemens.

R.-K. Basisschool "St. Henricus" te Elst.

Hoofd der School : J.N. Schoonderbeek

Leerkrachten : A. Hijmans en H.J.M. Wissink.

2. Inspectie.

Inspectie Elst van het kleuteronderwijs en van het lager onderwijs:
inspectrice J.E.H. Fortgens.

Inspectie Nijmegen van het kleuteronderwijs en van het lager onderwijs:
waarnemend inspecteur W. Schols.

Inspectie Wijchen van het kleuteronderwijs en van het lager onderwijs:
inspecteur W.S. van Dinter.

3. Test-instructeurs.

A. Cleiren, M. de Jongh, W. van Galen, J. Kuys, H. van Leent,
H. Migchielsen, M. Verhoef, J. Willems en B. Zandbergen.

4. Programma-instructeurs.

B. Gademann, G. van Hal, H. Kooiman, H. Lassche, P. Lem, C. Noordeloos,
C. Parren, M. Swets, S. Verwiël.

5. Assistenten en leerkrachten programma-ontwikkeling.

Assistenten : P. van Hoewijk, C. Noordeloos en Th. Rensen

Leerkrachten: J. van Liempt, J. van Wanroy, H. Bressers.

6. Adviseurs.

Drs. J. van Leeuwe, hoofd van de Research Technische Dienst.

Drs. M. Voeten, methodoloog.

7. Samensteller vragenlijst.

J.M. Berkelaar-Tomesen.

8. Assistent dataverwerking.

H. Feenstra.

9. Typiste.

A. van Loveren.

10. Vertaling.

C.J. Theunissen-Burke.

VRAGENFORMULIER TESTINSTRUCTEURS

instructeur _____ (1)
 school _____ (2)
 klas _____ (2)
 naam test _____ (3-4)
 datum _____ van _____ uur tot _____ uur (5-10)

Leerlingen afwezig? nee/ja^x

Zo ja, wie?

code nummer	naam
1	_____
2	_____
3	_____
4	_____
5	_____
6	_____
7	_____
8	_____
9	_____
10	_____ (11-12)

Leerlingen die de test in de maximaal beschikbare tijd niet af kregen?

Nee/ja^x

Zo ja, wie?

code nummer	naam	code nummer	naam
1	_____	11	_____
2	_____	12	_____
3	_____	13	_____
4	_____	14	_____
5	_____	15	_____
6	_____	16	_____
7	_____	17	_____
8	_____	18	_____
9	_____	19	_____
10	_____	20	_____ (13-14)

Waren er leerlingen die, voordat ze de test af hadden, en voor het verstrijken van de maximale test-tijd, genoodzaakt waren met het maken van de test te stoppen?

Ja/nee^x

Zo ja, wie?

code nummer	naam
1	_____
2	_____
3	_____
4	_____
5	_____
6	_____
7	_____ (15-16)

Hebt u opmerkingen over het testmateriaal? nee/ja^x (17)
Zo ja, welke?

Hebt u opmerkingen over het verloop van de testafname? nee/ja^x (18)
Zo ja, welke?

Heeft de leerkracht opmerkingen gemaakt over het testmateriaal of (19)
over de testafname? nee/ja^x
Zo ja, welke?

De aangegeven instructietijd was voor deze afname:

ruim voldoende / precies goed / te krap^x (20)

Wat is uw mening over de maximale afnameduur van de test (gezien de reacties van de leerlingen)?

ruim voldoende / kan net / te krap^x (21)

Hoeveel tijd kostte u: - het maken van afspraken?^{xx} _____ min.(22-24)
- de voorbereiding van de toetsafname? _____ min.(25-27)
- de toetsafname? _____ min.(28-30)
- de controle/codering van het werk? _____ min.(31-33)
- de reistijd? _____ min.(34-36)
- de overige bezigheden? _____ min.(37-39)

Omschrijving overige bezigheden: _____

^x Doorstrepen wat niet van toepassing is.

^{xx} Bij afname van meer toetsen de benodigde tijd gelijkelijk over de verschillende tests/klassen verdelen.

Bijlage 4.3.

Lessen uit het programma "Begrijpend lezen" van E. Bol, E. Köllen en J. Leistra (controleprogramma).

Les 1	blz. 8 t/m 12 bovenaan.
Les 2	blz. 12 t/m 16.
Les 3	blz. 25 t/m 28 middenin. tekst "Winnie de Pooh" i.p.v. tekst "Verkeersborden"
Les 4	blz. 28 middenin t/m 33. tekst "Verkeersborden" i.p.v. tekst "Winnie de Pooh"
Les 5	blz. 45 t/m 47 middenin.
Les 6	blz. 47 middenin t/m 48. meer teksten voor les 6.
Les 7	blz. 53 t/m 56.
Les 8	blz. 63 t/m 65.
Les 9	blz. 49 t/m 53 middenin.
Les 10	blz. 53 middenin t/m 56.

Concept draaiboek 3e klas "Begrijpend lezen" S.V.O. 0-275, Utrecht 1979.

VRAGENFORMULIER INSTRUCTEURS

instructeur _____ (1)
 school _____ (2)
 groep _____ (3)
 les _____ (4-5)
 datum _____ tijd _____ (6-11)

1. Waren er leerlingen afwezig? nee/ja^x

Zo ja, wie?

code nummer	naam
1	_____
2	_____
3	_____
4	_____
5	_____
6	_____

(12-13)

2. Waren er leerlingen die de les niet helemaal hebben kunnen bijwonen?

nee/ja^x

Zo ja, wie?

code nummer	naam
1	_____
2	_____
3	_____
4	_____

(14-15)

3. Hoeveel tijd heeft de les geduurd? Schrijf het aantal minuten op.

_____ minuten.

(16-17)

4. Denk je dat de meeste leerlingen het doel van de les hebben bereikt?

Ja/nee/weet ik niet^x

(18)

Indien de les meerdere doelen bevat:

Is doel 1 bereikt? Ja/nee/weet ik niet^x

(19)

Is doel 2 bereikt? Ja/nee/weet ik niet^x

(20)

Is doel 3 bereikt? Ja/nee/weet ik niet^x

(21)

5. De motivatie van de meeste leerlingen in deze les was:

zeer goed/goed/ruim voldoende/zwak/zeer zwak^x.

(22)

Wat is de belangrijkste reden waarom de leerlingen zo waren gemotiveerd?

6. Kwamen er tijdens de les ernstige fricties of storingen voor? (23)
Ja/nee*. Indien ja, beschrijf kort de aard ervan.

7. Heb je de les in de beschikbare tijd afgekregen? Ja/nee*. (24)
Indien nee, geef dan in de handleiding precies aan welke activiteiten of oefeningen niet zijn behandeld of uitgevoerd. Zet daarbij de opmerking: "niet behandeld of uitgevoerd".

8. Welke activiteiten of oefeningen verliepen volgens plan?
Zet in de handleiding naast de geplande activiteiten of oefeningen de opmerking "liep goed" als je van mening bent dat die activiteiten of oefeningen inderdaad goed verliepen.

9. Welke activiteiten of oefeningen verliepen niet volgens plan?
Zet in de handleiding naast de geplande activiteiten of oefeningen de opmerking "liep niet goed" als je van mening bent dat die activiteiten of oefeningen inderdaad niet goed verliepen. Schrijf er zo mogelijk bij waarom dat het geval was.

10. Hoeveel tijd kostte je:

de voorbereiding van deze les	_____	minuten	(25-27)
de uitvoering van deze les	_____	minuten	(28-30)
het invullen van het vragenform.	_____	minuten	(31-33)
de reistijd	_____	minuten	(34-36)
de overige bezigheden	_____	minuten	(37-39)

Omschrijving overige bezigheden: _____

* Doorstrepen wat niet van toepassing is.

EVALUATIE VAN HET HELE PROGRAMMA

Instructeur _____

Programma _____

School _____

1. Hoe belangrijk vind je het programma voor leerlingen van de vierde klas?^{*}

zeer belangrijk / belangrijk / niet zo belangrijk / onbelangrijk / zeer on-
belangrijk

Opmerkingen: _____

2. Wat is je mening over de moeilijkheidsgraad van het programma voor leerlingen van de vierde klas?^{*}

zeer moeilijk / moeilijk / niet moeilijk en niet makkelijk / gemakkelijk /
zeer gemakkelijk

Opmerkingen: _____

3. Wat is je mening over de structuur of opbouw van het programma?^{*}

zeer goed / goed / voldoende / zwak / zeer zwak

Opmerkingen: _____

^{*} Streep door wat niet van toepassing is.

4. De aanwijzingen voor de instructeur waren meestal^x:

zeer duidelijk / duidelijk / onduidelijk / zeer onduidelijk

zeer overzichtelijk / overzichtelijk / onoverzichtelijk / zeer onover-
zichtelijk

Opmerkingen: _____

5. Het leerlingenmateriaal was meestal^x:

zeer duidelijk / duidelijk / onduidelijk / zeer onduidelijk

zeer overzichtelijk / overzichtelijk / onoverzichtelijk / zeer onover-
zichtelijk

Opmerkingen: _____

6. De lessen die je gegeven hebt, waren nogal voorgestructureerd en gespecificeerd. Hoe prettig vond je het om deze lessen te geven^x?

zeer prettig / prettig / niet zo prettig / onprettig

Opmerkingen: _____

^x Streep door wat niet van toepassing is.

7. Je hebt de les gegeven aan leerlingen van de school te: _____
(invullen). Hoe prettig vond je het om deze leerlingen les te geven^x?
zeer prettig / prettig / niet zo prettig / onprettig

Opmerkingen: _____

8. De wekelijkse bijeenkomsten ter voorbereiding van de lessen waren meest-
al^x:

zeer nuttig / nuttig / niet zo nuttig / overbodig

Opmerkingen: _____

9. Schrijf hieronder welke op- en aanmerkingen je verder hebt over het pro-
gramma en de uitvoering daarvan.

S.V.P. ZO SNEL MOGELIJK RETOUR. BEDANKT.

FRCT 12

	Totale groep		Controle- scholen	Exp. scholen
	max.	constant		
<u>Voormeting</u>				
school	1 - 8	zonder 2, 5	6, 7, 8	1, 3, 4
aantal leerlingen	520	414	226	188
gemiddelde p	.61	.62	.63	.61
standaarddeviatie	7.1	7.1	6.8	7.4
KR-20	.86	.86	.85	.87
<u>Nameting</u>				
school	zonder 2	zonder 2, 5	6, 7, 8	1, 3, 4
aantal leerlingen	457	404	219	185
gemiddelde p	.67	.67	.69	.65
standaarddeviatie	7.2	7.2	6.8	7.5
KR-20	.85	.85	.84	.86
<u>Retentiemeting</u>				
school	zonder 2, 5	zonder 2, 5	6, 7, 8	1, 3, 4
aantal leerlingen	407	407	221	186
gemiddelde p	.70	.70	.70	.68
standaarddeviatie	6.2	6.2	6.1	6.3
KR-20	.84	.84	.84	.84
<u>Correlaties tussen meetmomenten</u>				
school	zonder 2	zonder 2, 5	6, 7, 8	1, 3, 4
aantal leerlingen	443	394	214	180
voormeting x nameting	.75	.75	.75	.75
school	zonder 2, 5	zonder 2, 5	6, 7, 8	1, 3, 4
aantal leerlingen	397	397	216	181
voormeting x retentiemeting	.78	.78	.78	.78
school	zonder 2, 5	zonder 2, 5	6, 7, 8	1, 3, 4
aantal leerlingen	393	393	210	183
nameting x retentiemeting	.75	.75	.75	.75

MCCT

	Totale groep		Controle- scholen	Exp. scholen
	max.	constant		
<u>Voormeting</u>				
school	1 - 8	zonder 2, 5	6, 7, 8	1, 3, 4
aantal leerlingen	513	410	224	186
gemiddelde p	.68	.69	.69	.69
standaarddeviatie	14.9	14.2	14.8	13.5
coëff. alpha over subtests	.85	.84	.84	.84
<u>Nameting</u>				
school	zonder 2	zonder 2, 5	6, 7, 8	1, 3, 4
aantal leerlingen	446	401	218	183
gemiddelde p	.72	.73	.73	.72
standaarddeviatie	14.4	14.3	13.9	14.6
coëff. alpha over subtests	.88	.88	.87	.89
<u>Retentiemeting</u>				
school	zonder 2, 5	zonder 2, 5	6, 7, 8	1, 3, 4
aantal leerlingen	410	410	222	188
gemiddelde p	.76	.76	.76	.76
standaarddeviatie	15.3	15.3	15.8	14.7
coëff. alpha over subtests	.90	.90	.90	.90
<u>Correlaties tussen meetmomenten</u>				
school	zonder 2	zonder 2, 5	6, 7, 8	1, 3, 4
aantal leerlingen	424	384	209	175
voormeting x nameting	.70	.71	.70	.73
school	zonder 2, 5	zonder 2, 5	6, 7, 8	1, 3, 4
aantal leerlingen	393	393	213	180
voormeting x retentiemeting	.73	.73	.65	.72
school	zonder 2, 5	zonder 2, 5	6, 7, 8	1, 3, 4
aantal leerlingen	395	395	212	183
nameting x retentiemeting	.78	.78	.82	.74

HOOFDGEDACHTETEST

	Totale groep		Controle scholen	Exp. scholen
	max.	constant		
<u>Voormeting</u>				
school	1 - 8	zonder 2, 5	6, 7, 8	1, 3, 4
aantal leerlingen	497	398	217	181
gemiddelde p	.50	.52	.53	.52
standaarddeviatie	7.0	7.0	7.1	6.9
KR-20	.89	.89	.89	.89
correlatie deel A1 x A2	.76	.74	.76	.72
idem, gecorrigeerd volgens S-B	.86	.85	.86	.84
<u>Nameting</u>				
school	zonder 2	zonder 2, 5	6, 7, 8	1, 3, 4
aantal leerlingen	456	403	212	191
gemiddelde p	.62	.64	.64	.63
standaarddeviatie	6.7	6.7	7.0	6.4
KR-20	.89	.89	.90	.87
correlatie deel B1 x B2	.74	.74	.79	.69
idem, gecorrigeerd volgens S-B	.85	.85	.88	.82
<u>Retentiemeting</u>				
school	zonder 2, 5	zonder 2, 5	6, 7, 8	1, 3, 4
aantal leerlingen	412	412	225	187
gemiddelde p	.67	.67	.67	.67
standaarddeviatie	6.6	6.6	6.7	6.4
KR-20	.89	.89	.89	.88
correlatie deel A1 x A2	.77	.77	.78	.75
idem, gecorrigeerd volgens S-B	.87	.87	.88	.86
<u>Correlaties tussen meetmomenten</u>				
school	zonder 2	zonder 2, 5	6, 7, 8	1, 3, 4
aantal leerlingen	427	383	203	180
voormeting x nameting	.75	.73	.75	.72
school	zonder 2, 5	zonder 2, 5	6, 7, 8	1, 3, 4
aantal leerlingen	387	387	211	176
voormeting x retentiemeting	.72	.72	.73	.72
school	zonder 2, 5	zonder 2, 5	6, 7, 8	1, 3, 4
aantal leerlingen	397	397	207	190
nameting x retentiemeting	.82	.82	.81	.83

BEDOELINGSTEST

	Totale groep		Controle- scholen	Exp. scholen
	max.	constant		
<u>Voormeting</u>				
school	1 - 8	zonder 2, 5	6, 7, 8	1, 3, 4
aantal leerlingen	487	393	220	173
gemiddelde p	.57	.59	.60	.58
standaarddeviatie	5.2	5.1	5.1	5.2
KR-20	.83	.83	.83	.83
correlatie deel A1 x A2	.68	.66	.63	.71
idem, gecorrigeerd volgens S-B	.81	.80	.77	.83
<u>Nameting</u>				
school	zonder 2	zonder 2, 5	6, 7, 8	1, 3, 4
aantal leerlingen	460	411	222	189
gemiddelde p	.65	.66	.66	.67
standaarddeviatie	5.2	5.0	5.2	4.9
KR-20	.84	.83	.84	.82
correlatie deel B1 x B2	.65	.63	.64	.62
idem, gecorrigeerd volgens S-B	.79	.77	.78	.77
<u>Retentiemeting</u>				
school	zonder 2, 5	zonder 2, 5	6, 7, 8	1, 3, 4
aantal leerlingen	413	413	227	186
gemiddelde p	.70	.70	.70	.70
standaarddeviatie	5.1	5.1	5.3	4.9
KR-20	.86	.86	.87	.84
correlatie deel A1 x A2	.70	.70	.73	.66
idem, gecorrigeerd volgens S-B	.82	.82	.84	.80
<u>Correlaties tussen meetmomenten</u>				
school	zonder 2	zonder 2, 5	6, 7, 8	1, 3, 4
aantal leerlingen	427	381	211	170
voormeting x nameting	.72	.71	.68	.74
school	zonder 2, 5	zonder 2, 5	6, 7, 8	1, 3, 4
aantal leerlingen	386	386	216	170
voormeting x retentiemeting	.74	.74	.68	.75
school	zonder 2, 5	zonder 2, 5	6, 7, 8	1, 3, 4
aantal leerlingen	404	404	219	185
nameting x retentiemeting	.75	.75	.75	.76

REDENEERTEST: TOTAAL

	Totale groep		Controle- scholen	Exp. scholen
	max.	constant		
<u>Voormeting</u>				
school	1 - 8	zonder 2, 5	6, 7, 8	1, 3, 4
aantal leerlingen	496	395	216	179
gemiddelde p	.46	.47	.47	.48
standaarddeviatie	6.9	7.2	7.1	7.2
KR-20	.83	.84	.84	.84
correlatie deel A1 x A2	.71	.72	.70	.74
idem, gecorrigeerd volgens S-B	.83	.84	.82	.85
<u>Nameting</u>				
school	zonder 2	zonder 2, 5	6, 7, 8	1, 3, 4
aantal leerlingen	467	415	223	192
gemiddelde p	.63	.64	.64	.65
standaarddeviatie	7.9	7.8	7.6	8.2
KR-20	.88	.88	.87	.89
correlatie deel B1 x B2	.75	.76	.75	.76
idem, gecorrigeerd volgens S-B	.86	.86	.86	.86
<u>Retentiemeting</u>				
school	zonder 2, 5	zonder 2, 5	6, 7, 8	1, 3, 4
aantal leerlingen	407	407	225	182
gemiddelde p	.68	.68	.67	.69
standaarddeviatie	8.2	8.2	8.0	8.5
KR-20	.90	.90	.89	.91
correlatie deel A1 x A2	.75	.75	.77	.74
idem, gecorrigeerd volgens S-B	.86	.86	.87	.85
<u>Correlaties tussen meetmomenten</u>				
school	zonder 2	zonder 2, 5	6, 7, 8	1, 3, 4
aantal leerlingen	432	385	209	176
voormeting x nameting	.69	.68	.66	.70
school	zonder 2, 5	zonder 2, 5	6, 7, 8	1, 3, 4
aantal leerlingen	377	377	209	168
voormeting x retentiemeting	.70	.70	.64	.69
school	zonder 2, 5	zonder 2, 5	6, 7, 8	1, 3, 4
aantal leerlingen	400	400	218	182
nameting x retentiemeting	.79	.79	.76	.82

REDENEERTEST: ZEKER

	Totale groep		Controle- scholen	Exp. scholen
	max.	constant		
<u>Voormeting</u>				
school	1 - 8	zonder 2, 5	6, 7, 8	1, 3, 4
aantal leerlingen	496	393	215	178
gemiddelde p	.50	.51	.51	.51
standaarddeviatie	4.9	5.0	5.2	4.8
KR-20	.80	.81	.82	.79
correlatie deel A1 x A2	.64	.65	.67	.63
idem, gecorrigeerd volgens S-B	.78	.79	.80	.77
<u>Nameting</u>				
school	zonder 2	zonder 2, 5	6, 7, 8	1, 3, 4
aantal leerlingen	463	415	223	192
gemiddelde p	.66	.67	.68	.67
standaarddeviatie	5.0	5.0	4.7	5.2
KR-20	.83	.83	.81	.85
correlatie deel B1 x B2	.69	.69	.64	.73
idem, gecorrigeerd volgens S-B	.82	.82	.78	.84
<u>Retentiemeting</u>				
school	zonder 2, 5	zonder 2, 5	6, 7, 8	1, 3, 4
aantal leerlingen	407	407	225	182
gemiddelde p	.70	.70	.70	.70
standaarddeviatie	5.2	5.2	5.0	5.5
KR-20	.86	.86	.84	.88
correlatie deel A1 x A2	.70	.70	.68	.73
idem, gecorrigeerd volgens S-B	.82	.82	.81	.84
<u>Correlaties tussen meetmomenten</u>				
school	zonder 2	zonder 2, 5	6, 7, 8	1, 3, 4
aantal leerlingen	430	385	209	176
voormeting x nameting	.55	.55	.52	.58
school	zonder 2, 5	zonder 2, 5	6, 7, 8	1, 3, 4
aantal leerlingen	377	377	209	168
voormeting x retentiemeting	.51	.51	.45	.58
school	zonder 2, 5	zonder 2, 5	6, 7, 8	1, 3, 4
aantal leerlingen	400	400	218	182
nameting x retentiemeting	.72	.72	.71	.74

REDENEERTEST: ONZEKER

	Totale groep		Controle- scholen	Exp. scholen
	max.	constant		
<u>Voormeting</u>				
school	1 - 8	zonder 2, 5	6, 7, 8	1, 3, 4
aantal leerlingen	496	393	216	177
gemiddelde p	.41	.42	.41	.44
standaarddeviatie	4.0	4.0	3.9	4.2
KR-20	.83	.82	.82	.83
correlatie deel A1 x A2	.71	.71	.69	.72
idem, gecorrigeerd volgens S-B	.83	.83	.82	.84
<u>Nameting</u>				
school	zonder 2	zonder 2, 5	6, 7, 8	1, 3, 4
aantal leerlingen	463	415	223	192
gemiddelde p	.59	.60	.58	.63
standaarddeviatie	4.1	4.0	4.0	4.0
KR-20	.83	.83	.83	.83
correlatie deel B1 x B2	.71	.71	.72	.71
idem, gecorrigeerd volgens S-B	.83	.83	.84	.83
<u>Retentiemeting</u>				
school	zonder 2, 5	zonder 2, 5	6, 7, 8	1, 3, 4
aantal leerlingen	407	407	225	182
gemiddelde p	.64	.64	.61	.68
standaarddeviatie	4.1	4.1	4.1	4.0
KR-20	.84	.84	.83	.84
correlatie deel A1 x A2	.71	.71	.74	.66
idem, gecorrigeerd volgens S-B	.83	.83	.85	.80
<u>Correlaties tussen meetmomenten</u>				
school	zonder 2	zonder 2, 5	6, 7, 8	1, 3, 4
aantal leerlingen	430	385	209	176
voormeting x nameting	.64	.63	.62	.64
school	zonder 2, 5	zonder 2, 5	6, 7, 8	1, 3, 4
aantal leerlingen	377	377	209	168
voormeting x retentiemeting	.60	.60	.59	.61
school	zonder 2, 5	zonder 2, 5	6, 7, 8	1, 3, 4
aantal leerlingen	400	400	218	182
nameting x retentiemeting	.75	.75	.72	.78

	FRCT 12		MCCT	
	Vorm A	Vorm B	Vorm A	Vorm B
aantal scholen/klassen	2/4	2/4	2/4	2/4
aantal leerlingen	103	98	97	97
gemiddelde p	.62	.65	.78	.79
standaarddeviatie	5.96	5.93	9.29	9.11
KR-20/coëfficiënt alpha	.79	.77	.80	.80
correlatie vorm A x vorm B	.74		.84	

	HGD		BED	
	vorm A	vorm B	vorm A	vorm B
aantal scholen/klassen	6	6	5	5
aantal leerlingen	186	186	127	127
gemiddelde p	.61	.60	.68	.68
standaarddeviatie	6.95	6.99	4.8	4.7
KR-20	.89	.89	.83	.82
correlatie vorm A x vorm B	.92		.85	

	RED					
	Totaal		Zeker		Onzeker	
	vorm A	vorm B	vorm A	vorm B	vorm A	vorm B
aantal scholen/klassen	6	6	6	6	6	6
aantal leerlingen	161	161	161	161	161	161
gemiddelde p	.65	.65	.73	.72	.53	.53
standaarddeviatie	7.5	7.4	4.8	4.7	4.5	4.4
KR-20	.87	.87	.84	.83	.85	.85
correlatie vorm A x vorm B	.90		.89		.88	

Bijlage 5.2. Correlatiematrices gebruikt bij de berekening van variantie-componenten van de ontwikkelde tests.

a. Gegevens voormeting; alle scholen; lijstgewijs weglaten van leerlingen.
(N = 383).

	1	2	3	4	5	6	7	8	9	10
1. Eén-Minuut-Test	-									
2. Woordenschattest	43	-								
3. Stilleestest	38	69	-							
4. Schriftelijke Opdrachten	49	76	64	-						
5. Hoofdgedachtetest	30	70	58	60	-					
6. Bedoelingtest	39	69	66	60	70	-				
7. Redeneertest	26	56	51	53	60	58	-			
8. Onzeker	24	57	52	52	60	56	73	-		
9. Zeker	17	32	30	33	36	37	82	22	-	
10. FRCT	48	78	62	73	61	59	47	49	26	-
11. MCCT	50	72	55	69	57	54	44	45	25	68

b. Gegevens voormeting; alle scholen; paarsgewijs weglaten van leerlingen.

	N(max)	1	2	3	4	5	6	7	8	9	10
1. Eén-Minuut-Test	519	-									
2. Woordenschattest	502	44	-								
3. Stilleestest	516	38	67	-							
4. Schriftelijke Opdrachten	515	51	77	62	-						
5. Hoofdgedachtetest	497	32	71	59	61	-					
6. Bedoelingtest	487	40	69	66	62	70	-				
7. Redeneertest	496	25	57	50	53	60	56	-			
8. Onzeker	494	24	69	51	50	60	55	72	-		
9. Zeker	494	15	34	29	33	35	34	82	20	-	
10. FRCT	519	47	76	60	74	59	61	48	48	28	-
11. MCCT	505	50	70	52	68	56	54	45	44	27	68

c. Gestandaardiseerde tests; voormeting; drie controlescholen.

	N(max)	1	2	3
1. Eén-Minuut-Test	227	-		
2. Woordenschattest	222	43	-	
3. Stilleestest	226	38	69	-
4. Schriftelijke Opdrachten	221	48	74	61

d. Correlaties tussen gestandaardiseerde tests en ontwikkelde tests; drie controlescholen.

	Voormeting (vorm A)				Nameting (vorm B)				Retentiemeting (vorm A)			
	1	2	3	4	1	2	3	4	1	2	3	4 ^x
Hoofdgedachtetest	28	67	52	55	39	66	51	56	38	69	53	61
Bedoelingtest	33	63	63	52	36	66	53	62	39	63	49	64
Redeneertest	25	55	46	54	32	65	51	62	33	65	46	61
Onzeker	22	55	48	48	23	61	53	56	25	61	49	55
Zeker	18	34	26	38	32	53	37	51	32	54	34	54
FRCT	46	74	59	70	42	71	59	72	42	69	50	68
MCCT	45	69	44	66	44	67	52	64	46	62	47	64

^x Betekenis van de cijfers 1, 2, 3 en 4 en N(max) als in tabel c.

De intercorrelaties van de ontwikkelde tests van zowel de voormeting, de nameting als de retentiemeting op de drie controlescholen zijn eveneens gebruikt bij de berekening van variantiecomponenten.

Bijlage 5.3. Verandering over de meetmomenten per voormetingsniveau (laag, midden, hoog). Controlescholen 6, 7, 8.

VOORMETING	Voormeting			Nameting			Retentiemeting			Groeps- grens
	\bar{x}	s	n	\bar{x}	s	n	\bar{x}	s	n	
LAAG										
FRCT	23.1	4.1	67	28.6	5.7	61	29.5	5.8	63	V<28
MCCT	52.1	9.5	72	60.1	14.1	67	62.2	15.7	70	V<65
HGD	7.1	2.5	70	12.1	5.2	67	14.0	5.8	67	V<12
BED	7.8	2.3	64	11.2	4.2	62	12.0	4.3	63	V<12
RED	11.5	2.0	62	21.1	5.4	60	21.5	6.2	61	V<15
Zeker	6.7	2.1	71	14.2	3.9	70	15.1	4.3	70	V<10
Onzeker	2.4	1.5	71	6.5	3.4	69	7.3	3.7	70	V< 5

VOORMETING	Voormeting			Nameting			Retentiemeting			Groeps- grens
	\bar{x}	s	n	\bar{x}	s	n	\bar{x}	s	n	
MIDDEN										
FRCT	30.9	1.9	77	33.2	5.4	74	34.9	3.7	74	27<V<35
MCCT	68.7	2.3	69	69.9	9.8	64	71.3	11.2	64	64<V<73
HGD	14.8	2.2	67	19.0	5.4	59	19.1	4.7	66	11<V<19
BED	14.1	1.4	71	15.6	3.6	68	17.1	4.3	70	11<V<17
RED	16.7	1.3	71	23.7	6.5	69	25.1	7.0	71	14<V<20
Zeker	11.4	1.1	65	15.4	4.5	62	15.6	5.2	64	9<V<14
Onzeker	6.1	0.8	65	9.6	3.2	64	9.2	3.2	64	4<V< 8

VOORMETING	Voormeting			Nameting			Retentiemeting			Groeps- grens
	\bar{x}	s	n	\bar{x}	s	n	\bar{x}	s	n	
HOOG										
FRCT	38.2	2.6	82	39.9	4.1	79	39.9	3.7	79	V>34
MCCT	79.9	5.1	81	81.1	7.8	78	84.4	8.4	79	V>72
HGD	22.9	2.8	80	24.1	4.3	77	24.9	3.9	78	V>18
BED	19.5	2.1	85	19.7	3.7	81	20.5	3.5	83	V>16
RED	26.0	5.6	83	31.4	6.2	80	32.4	6.5	78	V>19
Zeker	17.8	3.1	79	19.2	3.9	76	19.7	4.0	75	V>13
Onzeker	10.7	2.5	80	12.2	3.2	76	12.8	3.1	76	V> 7

Bijlage 5.4. Verandering over de meetmomenten per school.

SCHOOL 6	Voormeting			Nameting			Retentiemeting		
	\bar{x}	s	n	\bar{x}	s	n	\bar{x}	s	n
FRCT	31.1	8.2	61	34.2	7.9	60	34.2	6.9	61
MCCT	70.5	13.7	60	71.6	14.9	60	72.2	17.1	60
HGD	17.4	7.5	61	18.7	7.1	60	19.1	7.2	59
BED	15.2	5.5	60	16.5	5.2	62	17.2	5.5	60
RED	18.9	7.8	61	26.7	7.7	62	26.7	8.0	62
Zeker	13.0	5.4	61	16.9	4.6	62	17.3	4.9	62
Onzeker	5.9	3.8	61	9.7	4.7	62	9.4	4.7	62

SCHOOL 7	Voormeting			Nameting			Retentiemeting		
	\bar{x}	s	n	\bar{x}	s	n	\bar{x}	s	n
FRCT	32.3	6.0	86	35.3	6.1	83	37.2	4.5	85
MCCT	70.2	10.1	85	72.4	12.7	81	74.9	14.7	82
HGD	14.3	7.2	78	19.3	7.3	79	19.9	6.6	86
BED	13.6	4.9	80	16.3	4.8	85	17.1	4.8	87
RED	19.1	6.6	80	26.2	7.5	82	27.1	8.1	84
Zeker	12.8	5.1	79	17.1	4.5	82	17.6	5.0	84
Onzeker	6.3	4.3	80	9.1	3.7	82	9.6	3.8	84

SCHOOL 8	Voormeting			Nameting			Retentiemeting		
	\bar{x}	s	n	\bar{x}	s	n	\bar{x}	s	n
FRCT	30.3	6.3	79	33.7	6.6	77	33.6	6.4	76
MCCT	61.9	14.3	77	67.7	13.9	78	70.5	15.1	81
HGD	14.6	6.1	78	17.5	6.4	74	18.8	6.5	81
BED	14.4	5.1	80	14.6	5.3	76	16.3	5.6	81
RED	18.4	7.1	75	24.0	7.4	80	25.8	7.9	80
Zeker	11.0	4.9	75	14.8	4.7	79	15.7	4.8	79
Onzeker	7.5	3.5	75	9.3	3.8	79	10.2	4.0	79

Tabel a. Multivariate en univariate covariantie-analyse op de totaalscores van de vijf tests binnen de experimentele scholen.
Nameting.

Hoofdeffect scholen

F-ratio voor multivariate toets op gelijkheid van vectoren van gemiddelden is 4.0313.

Aantal vrijheidsgraden is 10 en 330 P minder dan 0.0001.

<u>variabele</u>	<u>gemid.kw.</u>	<u>univariate F</u>	<u>p minder dan</u>
FRCT	109.1854	5.5031	0.0049
MCCT	619.1444	6.5469	0.0019
HGD	2.7874	0.1678	0.8457
RED	299.7327	12.8620	0.0001
BED	73.0072	6.9256	0.0013
bij 2 en 169 vrijheidsgraden			

Hoofdeffect programma's

F-ratio voor multivariate toets op gelijkheid van vectoren van gemiddelden is 4.0420.

Aantal vrijheidsgraden is 15 en 456 P minder dan 0.0001.

<u>variabele</u>	<u>gemid.kw.</u>	<u>univariate F</u>	<u>p minder dan</u>
FRCT	2.9726	0.1498	0.9297
MCCT	46.4190	0.4908	0.6892
HGD	24.3464	1.4654	0.2259
RED	376.7590	16.1673	0.0001
BED	11.7460	1.1142	0.3449
bij 3 en 169 vrijheidsgraden			

Interactie-effect scholen x programma's

F-ratio voor multivariate toets op gelijkheid van vectoren van gemiddelden is 1.3741.

Aantal vrijheidsgraden is 30 en 662 P minder dan 0.0898.

<u>variabele</u>	<u>gemid.kw.</u>	<u>univariate F</u>	<u>p minder dan</u>
FRCT	15.2415	0.7682	0.5960
MCCT	117.8930	1.2466	0.2851
HGD	26.2214	1.5782	0.1563
RED	17.4446	0.7486	0.6114
BED	17.8941	1.6975	0.1245
bij 6 en 169 vrijheidsgraden			

Tabel b. Overzicht "gepoolde" binnen-cellen-varianties voor en na correctie van de factorscore in de na- en de retentiemeting.

test	nameting (B-vorm tests)		retentiemeting (A-vorm tests)	
	variantie	variantie na correctie	variantie	variantie na correctie
FRCT	58.19	19.84	40.53	13.79
MCCT	216.78	94.57	224.52	113.35
HGD	42.46	16.61	41.11	14.77
RED	62.39	23.30	71.89	31.23
BED	24.54	10.54	25.74	11.01
Factorscore voormeting	1.02	-	1.02	-

Tabel c. Geobserveerde celgemiddelden en standaarddeviaties van de afhankelijke variabelen uit de nameting en van de covariabele uit de voormeting. N = 182.

		<u>Gemiddelden</u>					Covariabele
<u>cel</u>		FRCT	MCCT	HGD	RED	BED	
<u>school</u>	<u>progr</u>						
1	BED	34.1	74.1	19.3	23.8	15.5	0.26
	HGD	33.4	72.7	19.4	24.4	16.8	0.35
	RED	34.4	67.8	18.4	29.4	17.4	0.27
	CONTR.	33.9	72.0	19.4	25.1	15.7	0.29
2	BED	32.4	70.2	18.9	27.0	16.7	-0.11
	HGD	32.7	70.7	19.3	25.4	15.7	-0.05
	RED	33.4	70.6	16.2	31.0	15.7	-0.04
	CONTR.	32.2	68.9	16.3	25.2	17.8	-0.07
3	BED	29.7	61.2	17.2	25.6	15.8	0.20
	HGD	32.3	67.8	19.8	22.1	13.8	0.20
	RED	30.2	64.2	19.4	29.5	15.0	0.10
	CONTR.	32.2	69.2	17.3	21.5	16.0	0.09

		<u>Standaarddeviaties</u>					Covariabele
<u>cel</u>		FRCT	MCCT	HGD	RED	BED	
<u>school</u>	<u>progr</u>						
1	BED	6.64	10.24	5.06	8.95	5.22	0.83
	HGD	7.50	15.20	5.50	7.17	3.70	0.92
	RED	6.31	19.62	6.79	8.60	4.23	0.93
	CONTR.	9.07	16.26	6.06	8.44	5.06	1.06
2	BED	7.69	14.33	6.02	7.30	4.62	0.99
	HGD	6.71	13.44	6.70	6.53	5.88	0.94
	RED	6.15	10.74	7.62	7.06	5.55	0.94
	CONTR.	7.69	9.07	6.87	6.68	3.37	0.96
3	BED	8.71	20.39	7.45	8.24	6.48	1.21
	HGD	6.21	16.45	6.97	8.32	6.06	1.12
	RED	10.27	18.28	6.53	7.56	4.79	1.20
	CONTR.	8.80	12.23	6.43	10.62	4.39	1.21

Tabel d. Geschatte celgemiddelden op basis van het hoofdeffectenmodel na uitpartialisering van de covariabele.
 ("cell means adjusted for covariate" in Multivariate).

Nameting

cel		FRCT	MCCT	HGD	RED	BED
school 1	Bedoeling	32.6	69.8	18.6	24.3	15.8
	Hoofdgedachte	32.8	70.4	19.1	22.6	15.1
	Redeneren	33.2	68.3	17.7	28.8	15.8
	Controle	33.1	70.4	17.6	22.9	16.3
school 2	Bedoeling	33.5	72.2	19.0	27.9	17.2
	Hoofdgedachte	33.7	72.9	19.5	26.3	16.5
	Redeneren	34.1	70.7	18.1	32.5	17.3
	Controle	34.0	72.9	17.9	26.6	17.7
school 3	Bedoeling	30.7	65.6	18.8	24.1	15.1
	Hoofdgedachte	30.9	66.2	19.2	22.4	14.4
	Redeneren	31.3	64.1	17.9	28.6	15.1
	Controle	31.2	66.2	17.7	22.8	15.6

Tabel e. Geschatte conditiegemiddelden voor de factor "programma's" in de nameting (schatting op basis van het hoofdeffectenmodel na uitpartialisering van de invloed van de covariabele).

conditie/test	FRCT	MCCT	HGD	RED	BED
programma Hoofdgedachte	32.4	69.8	19.3	23.8	15.3
programma Bedoeling	32.3	69.2	18.8	25.4	16.0
programma Rederen	32.9	67.7	17.9	30.0	16.1
Controleprogramma	32.7	69.8	17.7	24.1	16.5

Tabel f. Multivariate en univariate covariantie-analyse op de totaalscores van de vijf tests binnen de experimentele scholen.
Retentiemeting.

Hoofdeffect scholen

F-ratio voor multivariate toets op gelijkheid van vectoren van gemiddelden is 2.8221.

Aantal vrijheidsgraden is 10 en 330. P minder dan 0.0023.

<u>variabele</u>	<u>gemid.kw.</u>	<u>univariate F</u>	<u>p minder dan</u>
FRCT	28.7486	2.0851	0.1275
MCCT	129.5360	1.1428	0.3214
HGD	4.3499	0.2945	0.7453
RED	209.7820	6.7167	0.0016
BED	53.5161	4.8608	0.0089
bij 2 en 169 vrijheidsgraden			

Hoofdeffect programma's

F-ratio voor multivariate toets op gelijkheid van vectoren van gemiddelden is 1.9340.

Aantal vrijheidsgraden is 15 en 456. P minder dan 0.0186.

<u>variabele</u>	<u>gemid.kw.</u>	<u>univariate F</u>	<u>p minder dan</u>
FRCT	14.5642	1.0563	0.3693
MCCT	14.9712	0.1321	0.9409
HGD	18.5455	1.2558	0.2914
RED	149.8525	4.7979	0.0032
BED	9.8635	0.8959	0.4446
bij 3 en 169 vrijheidsgraden			

Interactie-effect scholen x programma's

F-ratio voor multivariate toets op gelijkheid van vectoren van gemiddelden is 0.8563.

Aantal vrijheidsgraden is 30 en 662. P minder dan 0.6887.

<u>variabele</u>	<u>gemid.kw.</u>	<u>univariate F</u>	<u>p minder dan</u>
FRCT	13.9311	1.0104	0.4203
MCCT	101.3970	0.8946	0.5003
HGD	20.3917	1.3808	0.2250
RED	18.6671	0.5977	0.7320
BED	6.7673	0.6147	0.7184
bij 6 en 169 vrijheidsgraden			

Tabel g. Geobserveerde celgemiddelden en standaarddeviaties van de afhankelijke variabelen uit de retentiemeting en van de covariabele uit de voormeting. N = 182.

		<u>Cemiddelden</u>					
cel		FRCT	MCCT	HGD	RED	BED	Covariabele
school	progr						
1	BED	35.0	76.3	21.1	27.3	17.1	0.26
	HGD	36.2	74.5	20.6	26.0	16.2	0.35
	RED	33.3	73.1	19.3	31.1	17.9	0.27
	CONTR.	33.9	70.1	19.9	26.7	16.9	0.29
2	BED	32.8	69.7	19.9	27.9	17.6	-0.11
	HGD	35.1	72.2	19.2	27.4	16.6	-0.05
	RED	34.2	72.2	17.6	29.9	16.8	-0.04
	CONTR.	34.2	73.4	17.7	28.9	17.9	-0.07
3	BED	34.0	71.5	18.8	25.6	17.0	0.20
	HGD	34.8	77.6	20.9	26.0	16.6	0.20
	RED	34.7	74.4	21.2	30.0	16.8	0.10
	CONTR.	35.8	75.2	18.9	24.7	15.6	0.09

		<u>Standaarddeviaties</u>					
cel		FRCT	MCCT	HGD	RED	BED	Covariabele
school	progr						
1	BED	5.79	15.01	4.88	8.19	4.66	0.83
	HGD	5.86	17.66	5.23	9.85	4.42	0.92
	RED	5.62	18.79	5.99	6.48	4.62	0.93
	CONTR.	7.55	20.74	7.31	9.57	5.20	1.06
2	BED	6.43	15.34	6.37	8.98	5.79	0.99
	HGD	6.94	11.57	7.00	7.63	5.63	0.94
	RED	7.11	12.55	7.20	8.12	5.05	0.94
	CONTR.	5.22	9.66	6.09	9.15	3.51	0.96
3	BED	6.24	13.94	7.67	8.32	5.33	1.21
	HGD	5.94	12.84	5.62	6.42	4.74	1.12
	RED	6.36	14.10	5.72	7.30	6.34	1.20
	CONTR.	6.81	13.67	7.34	10.34	5.42	1.21

Tabel h. Geschatte celgemiddelden op basis van het hoofdeffectenmodel na uitpartialisering van de covariabele.
 ("cell means adjusted for covariate" in Multivariate).

Retentiemeting

cel	FRCT	MCCT	HGD	RED	BED
school 1 Bedoeling	33.2	71.2	20.0	26.2	16.8
Hoofdgedachte	34.5	72.5	19.7	25.2	15.7
Redeneren	33.3	71.7	18.9	29.3	16.6
Controle	33.9	71.4	18.7	26.0	16.4
school 2 Bedoeling	34.5	73.3	20.3	29.2	19.3
Hoofdgedachte	35.8	74.6	20.0	28.3	17.3
Redeneren	34.6	73.9	19.2	32.4	18.2
Controle	35.2	73.5	18.9	29.1	18.0
school 3 Bedoeling	34.2	74.1	20.6	25.9	16.8
Hoofdgedachte	35.4	75.4	20.3	25.0	15.7
Redeneren	34.3	74.6	19.5	29.1	16.6
Controle	34.8	74.3	19.2	25.8	16.4

Tabel i. Geschatte conditiegemiddelden voor de factor "programma's in de retentiemeting (schatting op basis van het hoofdeffectenmodel na uitpartialisering van de invloed van de covariabele).

	FRCT	MCCT	HGD	RED	BED
programma Hoofdgedachte	35.2	74.1	20.0	26.2	16.2
programma Bedoeling	34.0	72.8	20.3	27.1	17.3
programma Redeneren	34.1	73.4	19.2	30.3	17.1
Controleprogramma	34.6	73.0	18.9	27.0	16.9

Tabel j. Gemiddelden en standaarddeviaties van tests, afgenomen in de voormeting in de experimentele en controlescholen.

experimentele scholen												controlescholen (4)			controlescholen (3)			
conditie	Bedoeling- programma			Hoofdgedachte- programma			Controle- programma			Redeneer- programma			geen programma			geen programma		
	N	gem.	st.d.	N	gem.	st.d.	N	gem.	st.d.	N	gem.	st.d.	N	gem.	st.d.	N	gem.	st.d.
Eén-Minuut-Test	46	69.4	12.69	48	71.0	13.62	45	70.5	15.89	48	69.4	12.89	276	66.6	14.71	227	68.1	14.77
Standard Progr. Matrices	46	37.3	7.38	46	37.8	8.49	47	36.9	8.95	48	36.9	10.28	277	38.1	9.09	228	38.5	9.14
Stilleestest	46	18.0	4.75	48	17.1	4.98	46	17.2	5.04	46	17.5	5.40	274	17.0	4.93	226	17.6	4.77
Woordenschat- test	45	38.5	9.80	44	38.4	10.67	44	37.4	9.94	42	37.9	10.25	271	38.2	9.97	222	39.3	9.60
Schriftelijke Opdrachten	47	19.1	6.38	48	19.5	6.10	47	18.3	7.00	46	19.3	6.39	270	19.2	6.48	221	20.2	5.96
Free Response Cloze Test	46	30.8	7.34	47	32.0	7.31	47	30.0	7.53	47	30.2	6.66	276	30.7	6.84	226	31.3	6.79
Multiple Choice Cloze Test	45	65.8	12.27	46	67.1	12.55	47	68.2	11.76	45	67.6	13.33	266	66.3	13.19	222	67.4	13.20
Hoofdgedachte- test	46	14.8	6.11	46	15.3	6.98	44	15.5	7.19	45	14.8	7.60	262	14.6	7.12	217	15.3	7.05
Redeneertest	45	19.4	6.85	44	18.2	6.81	45	19.8	7.44	44	19.5	7.74	263	18.5	6.96	216	18.8	7.10
Bedoelingtest	43	13.8	5.81	42	14.0	4.20	45	14.4	5.50	42	13.8	4.89	267	13.7	5.25	220	14.3	5.14
Factorscore	48	0.096	1.01	48	0.119	0.97	48	0.082	1.04	48	0.098	0.99	281	0.003	0.99	231	0.120	0.95

Tabel k. Multivariate en univariate enkelvoudige covariantie-analyse op de totaalscores van de vijf tests in de experimentele scholen en de controlescholen samen.

Nameting

F-ratio voor multivariate toets op gelijkheid van vectoren van gemiddelden is 4.3349.

Aantal vrijheidsgraden is 20 en 1434. P minder dan 0.0001.

<u>variabele</u>	<u>gemid.kw.</u>	<u>vrijheidsgr.</u>	<u>univariate F</u>	<u>p minder dan</u>
FRCT	72.5640	4 en 436	3.7798	0.0050
MCCT	79.9380	4 en 436	0.9246	0.4495
HGD	19.6431	4 en 436	1.0201	0.3966
RED	284.8735	4 en 436	11.3252	0.0001
BED	13.4448	4 en 436	1.1191	0.3469

Tabel l. Multivariate en univariate enkelvoudige covariantie-analyse op de totaalscores van de vijf tests in de experimentele scholen en de controlescholen samen.

Retentiemeting

F-ratio voor multivariate toets op gelijkheid van vectoren van gemiddelden is 1.9646.

Aantal vrijheidsgraden is 20 en 1271. P minder dan 0.0068.

<u>variabele</u>	<u>gemid.kw.</u>	<u>vrijheidsgr.</u>	<u>univariate F</u>	<u>p minder dan</u>
FRCT	22.8373	4 en 387	1.4180	0.2273
MCCT	21.8699	4 en 387	0.1887	0.9443
HGD	15.4469	4 en 387	0.8782	0.4770
RED	159.0978	4 en 387	5.0779	0.0006
BED	8.4882	4 en 387	0.6957	0.5954

Tabel m. Geschatte directe effecten van de taakgerichte programma's^x.

	Nameting			Retentiemeting		
	in experimentele scholen		t.o.v. controle-scholen	in experimentele scholen		t.o.v. controle-scholen
	t.o.v. controle-programma	t.o.v. alle andere programma's		t.o.v. controle-programma	t.o.v. alle andere programma's	
HGD	0.38 (0.04 - 0.72)	0.29 (0.09 - 0.49)	0.14 (-0.13 - 0.41)	0.28 (-0.06 - 0.62)	0.14 (-0.06 - 0.34)	0.16 (-0.11 - 0.43)
BED	-0.16 (-0.51 - 0.19)	0.01 (-0.19 - 0.21)	0.15 (-0.12 - 0.42)	0.11 (-0.23 - 0.45)	0.17 (-0.03 - 0.37)	0.17 (-0.10 - 0.44)
RED	1.21 (0.87 - 1.55)	1.15 (0.95 - 1.35)	0.89 (0.62 - 1.16)	0.59 (0.25 - 0.93)	0.63 (0.43 - 0.83)	0.72 (0.45 - 0.99)

^x De effecten zijn weergegeven in standaardafwijking-eenheden. Genomen is steeds de "gepoolde" binnen-cellen standaardafwijking van het betrokken design, na correctie voor de covariabele. Tussen haakjes 90% betrouwbaarheidsintervallen.

Lijst van gebruikte computerprogramma's.

- ITEMA : Itemanalyse op dichotome data. J. van Leeuwe, F. Gremmen & A. Bouts. Nijmegen: Pedagogisch Instituut, 1971.
- MULTIVARIANCE : Univariate and multivariate analysis of variance, covariance, regression and repeated measures. A fortran IV program. Version VI. J.D. Finn. Chicago, Ill.: National Educational Resources, 1977.
- SPSS : Statistical package for the social sciences. N.H. Nie et al. New York: McGraw-Hill Book Company, 1975².

CURRICULUM VITAE

Cornelis Antonius Jozef Aarnoutse werd op 18 maart 1939 te Gennep geboren. Na het behalen van de akte van bekwaamheid als volledig bevoegd onderwijzer en na vervulling van de militaire dienstplicht was hij vier jaar werkzaam in het gewoon lager onderwijs en drie jaar in het toenmalig kweekschoolonderwijs (Pedagogische Academie te Sittard). Naast zijn werk studeerde hij pedagogiek aan de Universiteit van Nijmegen. In 1965 slaagde hij cum laude voor het kandidaatsexamen en in 1968 cum laude voor het doctoraalexamen. In het laatst genoemde jaar werd hij benoemd tot wetenschappelijk medewerker van het Instituut voor Onderwijskunde van de Katholieke Universiteit te Nijmegen. Van 1972-1974 was hij (mede)projectleider van het onderzoek "Effectief vragen stellen" (S.V.O. 0237) en van 1974-1980 projectleider van drie onderzoeksprojecten op het gebied van lezen en leesonderwijs (S.V.O. 0298). Sedert 1980 werkt hij mee aan een onderzoek dat betrekking heeft op goede en zwakke lezers (S.V.O. 0523) en is hij betrokken bij het innovatie- en onderzoeksproject "Preventie van leesmoeilijkheden" (S.V.O. 0492). Gedurende een aantal jaren was hij lid van enkele schoolbesturen en van de bestuursraad van de Stichting voor de Leerplanontwikkeling. Ook was hij docent van de opleiding Pedagogiek MO-B van de Katholieke Leergangen te Tilburg. Zijn publicaties liggen op het terrein van begrijpend lezen en curriculumontwikkeling.

